Deliverable No. 12.4

Report on the clinical adaptation and optimization of the Oncosimulator models. *In silico* studies.

Grant Agreement No.:	270089
Deliverable No.:	D12.4
Deliverable Name:	Report on the clinical adaptation and optimization of the Oncosimulator models. *In silico* studies
Contractual Submission Date:	31/01/2014
Actual Submission Date:	17/02/2014

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

SEVENTH FRAMEWORK PROGRAMME

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	***p-medicine***
Project Full Name:	From data sharing and integration via VPH models to personalized medicine
Deliverable No.:	D12.4
Document name:	Report on the clinical adaptation and optimization of the Oncosimulator models. *In silico* studies.
Nature (R, P, D, O)[1]	R
Dissemination Level (PU, PP, RE, CO)[2]	PU
Version:	Final
Actual Submission Date:	17/02/2014
Editor: Institution: E-Mail:	Georgios Stamatakos ICCS-NTUA gestam@central.ntua.gr

ABSTRACT:

This document describes the clinical adaptation and optimization procedures applied to the three branches of the p-medicine Oncosimulator, namely the nephroblastoma (N) branch, the breast cancer (B) branch and the acute lymphoblastic leukemia (L) branch. For the N and L branches anonymized real human data sets have been utilized whereas for the B branch animal data has been utilized at this stage. The L branch is considerably supported by the utilization of machine learning techniques applied to the corresponding molecular data. Concrete adaptation results along with evaluation comments are provided for each separate branch of the p-medicine Oncosimulator.

KEYWORD LIST: *in silico* oncology, multiscale cancer modelling, oncosimulator, neprhroblastoma, vincristine, dactinomycin, breast cancer, bevacizumab, angiogenesis, neovasculature, acute lymphoblastic leukemia, ALL, prednisone, methotrexate

The research leading to these results has received funding from the European Community Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 270089.

The author is solely responsible for its content, it does not represent the opinion of the European Community and the Community is not responsible for any use that might be made of data appearing therein.

[1] **R**=Report, **P**=Prototype, **D**=Demonstrator, **O**=Other

[2] **PU**=Public, **PP**=Restricted to other programme participants (including the Commission Services), **RE**=Restricted to a group specified by the consortium (including the Commission Services), **CO**=Confidential, only for members of the consortium (including the Commission Services)

MODIFICATION CONTROL			
Version	Date	Status	Author
1.0	25/01/2014	Draft	Georgios Stamatakos, ICCS-NTUA
2.0	10/02/2014	Revision	Georgios Stamatakos, ICCS-NTUA
3.0	13/02/2014	Revision	Dimitra Dionysiou, ICCS-NTUA
4.0, Final	16/02/2014	Revision	Georgios Stamatakos, ICCS-NTUA

List of contributors

Eleni Georgiadi, ICCS-NTUA

Katerina Argyri, ICCS-NTUA

Eleni Kolokotroni, ICCS-NTUA

Eleftherios Ouzolunoglou, ICCS-NTUA

Norbert Graf, USAAR

Martin Stanulla, LUH

See Appendix A at the end of this document for a list of affiliation acronyms

Contents

Executive Summary……6
I. Introduction……7
N. Clinical adaptation and optimization of the nephroblastoma Oncosimulator using real clinical cases……8
N.1 The simulation model parameters……8
N.2 The effect of the Oncosimulator parameters on the virtual tumour characteristics……10
N.2.1 Growth rate – tumour doubling time……10
N.2.2 Cell constitution of virtual tumours……10
N2.3 Chemotherapy induced tumour volume reduction……11
N.3 Adaptation of model parameters to real clinical nephroblastoma tumour cases……11
N.3.1 General methodology……11
N.3.2 Clinical nephroblastoma cases data……11
N.3.3 Adaptation assumptions regarding the nephroblastoma cases……13
N.3.4 Adaptation results……15
N.4 REFERENCES……26
B. Report on the clinical adaptation and optimization of the breast cancer Oncosimulator models. *In silico* studies……27
B.1 Basic features and internal information flow of the vascular tumour growth model……27
B.2 Results of an exploratory set of parametric simulations for the case of an untreated tumour……31
B.3 Local fitting of the model: reproducing a series of *in vivo* experiments in mice treated with bevacizumab……38
B.4 REFERENCES……43
L. Report on the clinical adaptation and optimization of the acute lymphoblastic leukemia (ALL) Oncosimulator models. *In silico* studies……46
L.1 ALL: A preliminary study exploiting clinical trial data for the clinical adaptation of the core model……46
L.1.1 Introduction……46
L1.2 Clinical data……47
L1.3 The cytokinetic model……48
L.1.3.1 ALL and the cancer stem cell theory……48
L.1.3.2 Cell kinetics of ALL……49
L.1.3.4 Modelling of cancer growth and response to chemotherapy……52
L.1.3.5 Modelling of the action mechanisms of the drugs considered……53
L.1.3.5.1 Methotrexate induced cell loss……53
L.1.3.5.2 Prednisone induced cell loss……54
L.1.4 A preliminary clinical adaptation study……54
L.1.4.1 Adaptation methodology……55
L.1.4.2 Results……56
L.1.5 REFERENCES……57
L.2 Supporting the prediction of the ALL Oncosimulator parameter values by applying machine learning methods to gene expression data……60
L.2.1 Introduction……60
L.2.2 Pathway-based gene expression analysis……61
L.2.3 Supporting the prediction of the ALL Oncosimulator cell kill rate parameter value……62

L.2.3.1 Description of the available data……62
L.2.3.2 Data preprocessing, missing value imputation and pathway level aggregation……..63
L.2.3.2.1 Collapsing gene expression from probe level to gene level……..63
L.2.3.2.2 Gene filtering……..63
L.2.3.2.3 Missing value imputation……64
L.2.3.2.4 Gene-set / pathway level gene expression aggregation……64
L.2.3.3 Training the machine learning model and cross-validation results……65
L.2.3.3.1 Problem definition and machine learning algorithms tested……65
L.2.3.3.2 Cross-validation results using the entire feature set……..65
L.2.4 Conclusions……67
L.2.5 REFERENCES……67
C. Overall conclusions……69
Appendix A - Abbreviations and Acronyms……..70

Executive Summary

This deliverable outlines the clinical adaptation and optimization procedures regarding the three branches of the p-medicine Oncosimulator, namely the nephroblastoma (N) branch, the breast cancer (B) branch and the acute lymphoblastic leukemia (L) branch.

For the N branch the simulation model parameters are presented in a tabular format. Subsequently, the effect of the parameter values on the virtual tumour characteristics is quantified. Special emphasis is put on the tumour growth, the tumour doubling time, the cellular constitution of virtual tumours and the chemotherapy induced tumour volume reduction. The model parameters are then adapted to anonymized real clinical nephroblastoma tumour cases. A general methodology is presented followed by an outline of real nephroblastoma data sets and a number of adaptation assumptions. Adaptation results are presented for the real nephroblastoma cases considered and optimization aspects are briefly discussed.

For the B branch the basic features and the internal information flow of the vascular tumour growth model are delineated. Results for an exploratory set of parametric simulations for the case of an untreated tumour are presented. Subsequently, a local fitting of the model to animal data is undertaken. A series of *in vivo* experiments in mice treated with bevacizumab is reproduced *in silico*.

For the L branch a preliminary study exploiting clinical trial data for the clinical adaptation of the core model is presented. The clinical data made available is outlined. The cytokinetic model developed and adopted is delineated. The modelling approach of cancer growth and response to chemotherapy is described. Special emphasis is put on the modelling of the action mechanisms of the drugs considered (methotrexate and prednisone). A preliminary clinical adaptation study is presented. The adaptation methodology is followed by the corresponding results. The prediction of the L branch of the Oncosimulator parameter values is considerably supported by applying machine learning methods to gene expression data. A comparative exposition of several pathway based gene expression analysis methods is provided. A method for supporting the prediction of the L branch of the Oncosimulator cell kill rate parameter value is outlined. The available data is described and approaches to the data preprocessing, missing value imputation and pathway level aggregation are presented. The latter refer to collapsing gene expression from probe to gene level, gene filtering, missing value imputation and gene-set / pathway level gene expression aggregation. Subsequently, the machine learning model developed is trained and cross-validation results are presented. A discussion of the approach complements the presentation of the clinical adaptation and optimization of the L branch branch of the p-medicine Oncosimulator.

An overall discussion is included at the end of the document.

I. Introduction

In view of the envisaged clinical translation of the p-medicine Oncosimulator branches i.e. the nephroblastoma (N) branch, the breast cancer (B) branch and the acute lymphoblastic leukemia (ALL) branch, a clinical adaptation and optimization procedure of the constituent multiscale models has to be undertaken. Within this framework the present document is structured as follows.

Chapter N presents the simulation model parameters in a tabular format. Subsequently, the effect of the parameter values on the virtual tumour characteristics is quantified. Special emphasis is put on the tumour growth, the tumour doubling time, the cellular constitution of virtual tumours and the chemotherapy induced tumour volume reduction. The model parameters are then adapted to anonymized real clinical nephroblastoma tumour cases. A general methodology is presented followed by an outline of real nephroblastoma data sets and a number of adaptation assumptions. Adaptation results are presented for the real nephroblastoma cases considered and optimization aspects are briefly discussed.

Chapter B presents the basic features and the internal information flow of the vascular tumour growth model. Results for an exploratory set of parametric simulations for the case of an untreated tumour are presented. Subsequently, a local fitting of the model to animal data is undertaken. A series of *in vivo* experiments in mice treated with bevacizumab is reproduced *in silico.* The clinical data made available is outlined. The cytokinetic model developed and adopted is delineated. The modelling approach of cancer growth and response to chemotherapy is described. Special emphasis is put on the modelling of the action mechanisms of the drugs considered (methotrexate and prednisone). A preliminary clinical adaptation study is presented. The adaptation methodology is followed by pertinent results.

Chapter L is divided into two sub-chapters. Sub-chapter L.1 presents a preliminary study exploiting clinical trial data for the clinical adaptation of the core model. The clinical data made available are outlined. The cytokinetic model developed and adopted is delineated. The modelling approach of cancer growth and response to chemotherapy is described. Special emphasis is put on the modelling of the action mechanisms of the drugs considered (methotrexate and prednisone). A preliminary clinical adaptation study is presented. The adaptation methodology is followed by the corresponding results. Sub-chapter L.2 deals with the application of machine learning techniques on gene expression data in order to support the estimation of the parameter values of the L branch of the Oncosimulator. A comparative exposition of several pathway based gene expression analysis methods is provided. A method for supporting the prediction of the L branch of the Oncosimulator cell kill rate parameter value is outlined. The available data is described and approaches to the data preprocessing, missing value imputation and pathway level aggregation are presented. These refer to collapsing gene expression from probe to gene level, gene filtering, missing value imputation and gene-set / pathway level gene expression aggregation. Subsequently, the machine learning model developed is trained and cross-validation results are presented. A discussion of the approach complements the presentation of the clinical adaptation and optimization of the L branch branch of the p-medicine Oncosimulator.

At the end of the document an overall discussion is included.

N. Clinical adaptation and optimization of the nephroblastoma Oncosimulator using real clinical cases.

N.1 The simulation model parameters.

The nephroblastoma Oncosimulator developed by the *In Silico* Oncology Group (ISOG), National Technical University of Athens (NTUA) and serving as the core cancer multiscale simulation model of p-medicine is a predominantly discrete, clinically-oriented multiscale cancer model of solid tumour free growth and response to chemotherapy [N.1-N.4]. Several complex mechanisms involved in cancer such as proliferation, quiescence, differentiation and death (normal and chemo-induced) are incorporated in the model (Fig. N.1). Four basic cell categories i.e. stem, limited mitotic potential (LIMP), differentiated and dead cells are considered. The transition rules between them (Fig. N.1) are described by the tumour dynamic parameters of the model (Table N.I). Other parameters are used to model the initial shape and volume of the tumour, the tumour density and the chemotherapeutic scheme.

TABLE N.I: Definition of the model parameters describing tumour dynamics

Symbol (units)	Definition
T_c (h)	Cell cycle duration
T_{G0} (h)	G0 (dormant phase) duration, i.e. time interval before a dormant cell dies through necrosis
T_N (h)	Time needed for necrosis to be completed and its lysis products to be eliminated from the tumour
T_A (h)	Time needed for apoptosis to be completed and its products to be eliminated from the tumour
R_A (h^{-1})	Apoptosis rate of living stem and LIMP tumour cells (fraction of non-differentiated cells dying through apoptosis per hour)
R_{ADiff} (h^{-1})	Apoptosis rate of differentiated tumour cells per hour
R_{NDiff} (h^{-1})	Necrosis rate of differentiated tumour cells per hour
P_{G0toG1}	The fraction of stem or LIMP cells having just left the G0 compartment that re-enter the cell cycle
N_{LIMP}	The maximum number of mitoses that a LIMP cell can perform before becoming terminally differentiated
P_{sym}	Fraction of stem cells that perform symmetric division.
P_{sleep}	Fraction of cells that enter G0 phase following mitosis
CKR_{VCR}	Cell kill rate for the specific vincristine dose
CKR_{ACT}	Cell kill rate for the specific actinomycin-D dose
CKR_{TOTAL}*	Combined cell kill rate of the two drugs (dependent parameter)

TABLE N.II: Miscellaneous model parameters (unrelated to tumour dynamics)

Other model parameters			
Input_image	The initial tumour volume area defined by the annotated imaging data of the patient at time of diagnosis	**.raw	Imaging data
NBC	Number of biological cells typically contained within a geometrical cell (GC) of the discretizing mesh (assigned in relation to the GC volume)	106 (for GC volume=1mm^3)	Based on typical tumour cell densities. [12]
Margin percent	Acceptable temporary over-loading or under-loading of each geometrical cell as a fraction of unity	0.1	Based on exploratory runs
T_{init} (h)	Time interval between the pre-treatment imaging data acquisition and the first drug administration	-	Clinical data or exploratory simulation data
T_{pt_scan} (h)	Time interval between the last drug administration and the post-treatment imaging data acquisition	-	Clinical data or exploratory simulation data
$T_{admin,i}$	Drug administration instants	-	Treatment protocol/clinical data

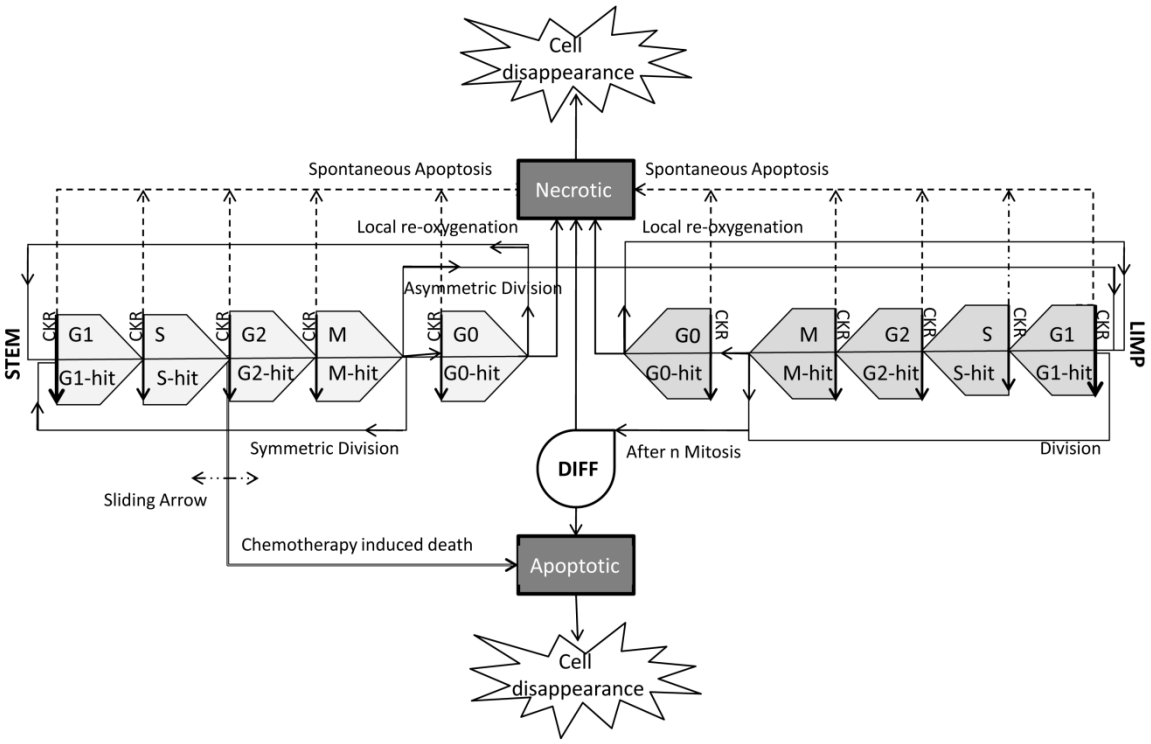


Fig. N.1: Generic cytokinetic model for tumour response to chemotherapy [N.2].
LIMP: limited mitotic potential cells, DIFF: terminally differentiated cells, G1: gap1 phase, S: DNA synthesis phase, G2: gap2 phase, M: mitosis phase, G0: dormant phase, Hit: Cells lethally hit by chemotherapy, CKR: Cell kill rate of the chemotherapeutic drugs. The sliding arrow indicates the point of the cell cycle at which cells are led to apoptosis and depends on the drugs mechanisms of action.

N.2 The effect of the Oncosimulator parameters on the virtual tumour characteristics.

A thorough sensitivity analysis of the model parameters has been performed [N.2, N.3] leading to a sorting of the parameters according to the magnitude of their effect on virtual tumour characteristics and has unveiled the relative importance of the corresponding biological mechanisms

N.2.1 Growth rate – Tumour doubling time

The virtual tumour growth rate depends only on the population of stem cells which preserve the tumour population. The growth fraction of a biologically accepted tumour is defined according to the equation [N.5]:

$$\left(1+P_{sym}\right)\left(1-P_{sleep}+P_{sleep}\frac{P_{G0toG1}/T_{G0}}{R_A+1/T_{G0}}\right)e^{-R_A T_C}\geq 1 \qquad \text{Equation N.1}$$

As depicted in eq.N.1 and based on the sensitivity analysis of the model [N.2, N.3], the most determinant model parameters on the growth fraction of virtual tumours are the following:

- Cell cycle duration (Tc)
- The fraction of stem cells that perform symmetric division (P_{sym}).
- The fraction of cells that enter G0 phase following mitosis (P_{sleep})
- Apoptosis rate of living stem and LIMP tumour cells (fraction of non-differentiated cells dying through apoptosis per hour) (R_A).

Other parameters that play a role in the value of the growth fraction of virtual tumour modeled, but with significantly reduced impact on it, are:

- the fraction of stem or LIMP cells having just left the G0 compartment that re-enter the cell cycle (P_{G0toG1})
- G0 (dormant phase) duration (T_{G0})

Decreasing the parameters T_c, P_{sleep}, R_A and T_{G0} and/or increasing the parameters P_{sym}, P_{G0toG1} lead to virtual tumours with higher growth fractions and lower doubling times i.e. more aggressive tumours.

N.2.2 Cell constitution of virtual tumours

The percentage of proliferating stem cells in the total population increases for a higher fraction of stem cells that perform symmetric division (P_{sym}), prolonged cell cycle duration (T_c) and lower loss rate due to spontaneous apoptosis (R_A).

The constitution of a tumour in terms of proliferating LIMP cells depends on two competitive mechanisms: the quantity of the producer (stem cells) of the LIMP population (P_{sym}) and the rate of production ($1-P_{sym}$). The duration of the cell cycle (T_c) and the loss rate (R_A) also affect the LIMP proliferating population in a way similar to the one they affect the stem cell proliferating population.

The fraction of the dormant stem and LIMP cells depends on the rate of quiescence (P_{sleep}), the rate of recycling (P_{G0toG1}) and the remaining time in the dormant state before entering necrosis.

The relative percentages of stem and committed progenitor cells are defined by the maximum number of mitoses that a LIMP cell can perform before becoming terminally differentiated (N_{LIMP}).
The differentiated population increases for a high percentage of LIMP cells and low rates of loss due to necrocis ($R_{NDiff)}$ and apoptosis (R_{ADiff}).

The fraction of dead cells in the total population increases if the loss rates (R_A, R_{Adiff}, R_{NDiff}) of the other populations increase or/and the time duration of the death products in the tumour bulk is prolonged (T_N, T_A).

N2.3 Chemotherapy induced tumour volume reduction

The chemotherapy induced volume reduction depends upon the growth rate of the virtual tumour and the cell kill rate of the chemotherapeutic drugs in use. Subsequently the model parameters that play an important role in the simulated tumour volume reduction are the T_c, P_{sym}, P_{sleep}, R_A and CKR_{TOTAL}. More aggressive tumours have a worse response to chemotherapy due to the fast repopulation of cancer cells. Increased effectiveness of chemotherapeutic drugs, leads to better therapeutic outcome.

N.3 Adaptation of model parameters to real clinical nephroblastoma tumour cases.

N.3.1 General methodology

First of all the miscellaneous model parameters (Table N.II) are adapted on a given patient. The real individualized imaging data is used to initialize the tumour region and the clinical data is used to set the chemotherapeutic scheme that has been administered to the specific patient. Subsequently, the model parameters describing the free growth tumour dynamics (Table N.I) are to be adapted. Primarily the model parameters that predominantly determine the tumour rate (as summarized in section N.2.1) should be adapted. In case that histological data of the patient is available and provides information on the (initial or/and final) tumour cell constitution, further parameters (e.g. cell cycle phase durations) that define the population fractions of cell subclasses have also to be adapted accordingly (see section N.2.2) .

Finally the therapy related tumour dynamics parameters, defining the effect of the chemotherapeutic drugs, are to be adapted based on the chemotherapy induced tumour volume reduction.

N.3.2 Clinical nephroblastoma cases data

In the framework of the p-Medicine project, 17 nephroblastoma tumour cases have been collected in the context of SIOP 2001/GPOH trial and analyzed in this study. The histological data of the patients are presented in Table N.III.

TABLE N.III: The histological data of nephroblastoma tumour cases selected in the context of SIOP 2001/GPOH trial

SIOP Number	Histology			Subpopulation			Histological type
	Regression/ Necrosis left	Regression/ Necrosis right	Metastasis	Blastemal	Ephitelial	Stromal	Staging
11351	30%		None	85%	10%	5%	blastemal predominant, high risk
11570		20%	None				mixed type, intermediate risk
11590		15%	None				mixed type with focal anaplasia, intermediate risk
11627		40%	None	5%	60%	35%	mixed type, intermediate malignancy
11628		30%	None	5%	25%	70%	stromal type, intermediate malignancy
11639	50%		None	0%	90%	10%	regressive type, intermediate malignancy
11803	<65%		None				stromal type, intermediate malignancy
11537		<65%	None		30%	70%	stromal type, internediate malignancy
11714	<65%		None	60%	25%	15%	mixed type, intermediate malignancy.
11733	<65%			100%			blastemal type, high malignancy
11736		<25%	None				mixed type, intermediate malignancy.
11813		<65%	None				mixed type, intermediate malignancy.
11823	90%		None	70%	20%	10%	regressive type, intermediate malignancy.
11845	80%		Lymphnodes				diffuse anaplasia, high malignancy,
11862	5%		None				epithelial type, intermediate malignancy.
11873		<65%	None	20%	40%	40%	mixed type, intermediate malignancy.
11881		80%	None				regressive type, intermediate malignancy.

*Bilateral nephroblastoma. **Not known.

As depicted in the above table (table N.III) 2 cases where blastemal predominant, 7 mixed type, 3 stromal, 3 regressive, 1 diffuse anaplasia and 1 epithelial nephroblastoma

TABLE N.IV: Clinical data of nephroblastoma tumour cases selected in the context of SIOP 2001/GPOH trial

SIOP Number	Date of diagnosis	Start Chemo-therapy	Cycle I (ACT:µg, VCR:mg)	Cycle II (VCR: mg)	Cycle III (ACT: µg, VCR: mg)	Cycle IV (VCR: mg)
11351	9/20/2005	9/29/2005	ACT: 931 VCR: 1,2	VCR: 1,2	ACT: 931 VCR: 1,2	VCR: 1,2
11570	8/30/2006	9/6/2006	ACT: 570 VCR: 0,85	VCR: 0,85	ACT: 570 VCR: 0,85	VCR: 0,85
11590	11/10/2006	11/10/2006	ACT: 900 VCR: 1,2	VCR: 1,2	ACT: 900 VCR: 1,2	VCR: 1,2
11627	2/5/2007	2/8/2007	ACT: 610 VCR: 0,9	VCR: 0,9	ACT: 610 VCR: 0,9	VCR: 0,9
11628	2/5/2007	2/9/2007	ACT: 880 VCR: 1,2	VCR: 1,2	ACT:880 VCR: 1,2	VCR: 1,2
11639	3/21/2007	3/23/2007	ACT: 740 VCR: 1	VCR: 1	ACT:740 VCR: 1	VCR: 1
11803	7/18/2008	7/18/2008	ACT: 605 VCR: 0,89	VCR: 0,89	ACT: 605 VCR: 0,89	VCR: 0,89
11537	6/13/2006	6/15/2006	ACT: 870 VCR: 1,2	VCR: 1,2	ACT: 870 VCR: 1,2	VCR: 1,2
11714	10/24/2007	10/25/2007	ACT: 350 VCR: 0,49	VCR: 0,49	ACT: 350 VCR: 0,49	VCR: 0,49
11733	12/6/2007	12/10/2007	ACT: 730 VCR: 1	VCR: 1	ACT: 730 VCR: 1	VCR: 1
11736	12/18/2007	12/19/2007	ACT: 310 VCR: 0,47	VCR: 0,47	ACT: 300 VCR: 0,47	VCR: 0,47
11813	8/21/2008	8/23/2008	ACT: 850 VCR:1,15	VCR:1,15	ACT: 850 VCR:1,15	VCR:1,15
11823	9/23/2008	9/25/2008	ACT: 1224 VCR:1,53	VCR:1,53	ACT: 1224 VCR:1,53	VCR:1,53
11845	11/25/2008	11/25/2008	ACT: 1200 VCR:1,50	VCR:1,50	ACT: 1200 VCR:1,50	VCR:1,50
11862	1/14/2009	1/19/2009	ACT: 267 VCR: 0,42	V: 0,42	ACT: 267 VCR: 0,42	VCR: 0,42
11873	1/30/2009	2/4/2009	ACT: 810 VCR: 1,1	VCR: 1,1	ACT: 810 VCR: 1,1	VCR: 1,1
11881	3/10/2009	3/11/2009	ACT: 399 VCR: 0,55	VCR: 0,82	ACT: 581 VCR: 0,82	VCR: 0,82

The anonymized imaging and clinical data has been collected by the clinicians in the framework of the SIOP 2001/GPOH trial and selected to be modelled by the 'Oncosimulator'. The clinical data of the nephroblastoma cases describing the chemotherapeutic scheme administered are presented in table N.IV.The initial and final virtual tumours have been spatiotemporally initialized based on MRI images of the clinical tumours collected at two time instants before the start of chemotherapy and after its completion.

N.3.3 Adaptation assumptions regarding the nephroblastoma cases

The three histological components of nephroblastoma (blastemal, epithelial and stromal) have different proliferating potential.The nuclear antigen Ki-67 is associated with cell proliferation, as it is present throughout the cell cycle and absent in resting cells [N.6] and is studied as a tumour

proliferation marker. At a study [N.6] the marker ki-67 SI was evaluated in the three components of nephroblastoma and was found to possess the highest values in blastemal and epithelial components. The median ki-67 SI was determined to be 33% in blastemal cells, 10% in stromal cells and 29.5% in epithelial cells. The study included postchemotherapy nephroblastomas, treated with neoadjuvant chemotherapy before nephrectomy. These values will be serve as gross initial assumptions for the adaptation of the growth fraction of the final tumour.

For mixed type nephroblastoma , the growth fraction will be evaluated according to the Equation N.2:

$$GF_{TOTAL} = p_{BLASTEMAL} * GF_{BLASTEMAL} + p_{EPITHELIAL} * GF_{EPITHELIAL} + p_{STROMAL} * GF_{STROMAL}$$
(Equation N.2)

Where:
GF: Growth fraction, p: percentage of the component
In case the histological subpopulation is not defined, the percentages of the three components are considered equal (even). For anaplastic nephroblastomas, the median ki-67 SI in the blastemal component was 70%, 24% in stromal and 43% in epithelial cells.For regressive type nephroblastomas a median ki-67 Si of 9.5 % is considered. According to the above assumptions, the calculated growth fraction of the final tumour to be adapted for the 17 cases, is noted in table N.V.

Another assumption for the adaptation of the nephroblastoma cases is a doubling time lying in the range of 11–40 days [N.7-N.11].

The imaging data based volume reduction is calculated for each nephroblastoma case (table N.V) and is subsequently used for the clinical adaptation procedure.

TABLE N.V: The chemotherapy induced tumour volume reduction for the nephroblastoma cases under study as calculated according to their imaging data before and after therapy.

SIOP Number	Relative clinical tumour volume reduction (%)	Growth fraction of the final clinical tumour (%)
11351	-9	33
11570	35	24
11590	63	24
11627	80	18
11628	82	10
11639	71	10
11803	39	10
11537	32	10
11714	2	27
11733	73	33
11736	25	24
11813	44	24
11823	70	10
11845	34	58
11862	71	30
11873	41	22
11881	73	10

N.3.4 Adaptation results

The acceptable temporary over-loading or under-loading of each geometrical cell as a fraction of unity is considered 0.1 for all the cases. All the cases undergo the chemotherapeutic treatment for Wilms tumour preoperative chemotherapy as defined by the SIOP trial treatment protocol and depicted in Fig. N.2.

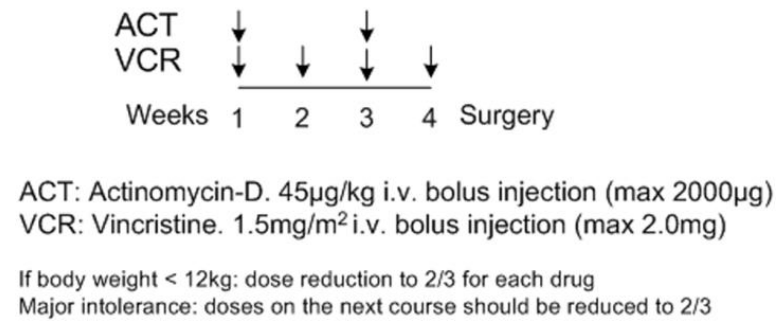


Fig. N.2: The simulated Wilms tumour preoperative chemotherapy treatment protocol of the SIOP/GPOH clinical trial.

The time interval between the pre-treatment imaging data acquisition and the first drug administration, and the time interval between the last drug administration and the post-treatment imaging data acquisition per case are defined by the available clinical data and presented in Table N.VI.

The adapted values of the model parameters are presented in Table VII. The resultant initial and final virtual tumour characteristics as well as the resultant tumour reduction, are given in Table VIII. In Fig. 3 the time evolution of the simulated tumour volumes for the 17 cases are depicted.

TABLE N. VI: The time interval between the pre-treatment imaging data acquisition and the first drug administration, and the time interval between the last drug administration and the post-treatment imaging data acquisition for 17 cases of nephroblastoma

Case	T_{init} (h)	T_{pt_scan} (h)
11351	144	0
11570	168	120
11590	72	96
11627	72	24
11628	24	144
11639	48	96
11803	144	240
11537	144	0
11714	24	312
11733	72	96
11736	24	120
11813	48	96
11823	48	96
11845	96	48
11862	120	216
11873	120	192
11881	24	120

TABLE N.VII: Values of dynamic model parameters assigned for the implementation of virtual tumours adapted to 17 clinical nephroblastoma cases.

Symbol (units)	**Adapted values**								
Case No	**11351**	**11570**	**11590**	**11627**	**11628**	**11639**	**11803**	**11537**	**11714**
T_c (h)	24	23	25	23	23	23	23	23	23
T_{G0} (h)	96	120	96	96	96	96	96	96	96
T_N (h)	20	20	20	20	20	20	20	20	20
T_A (h)	6	6	6	6	6	6	6	6	6
R_A (h^{-1})	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
R_{ADiff} (h^{-1})	0.045	0.02	0.015	0.015	0.005	0.005	0.003	0.003	0.018
R_{NDiff} (h^{-1})	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001
P_{G0toG1}	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
N_{LIMP}	7	7	7	7	7	7	7	7	7
P_{sym}	0.50	0.51	0.49	0.5	0.45	0.44	0.46	0.44	0.49
P_{sleep}	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
CKR_{VCR}	0.28	0.362	0.334	0.5	0.43	0.31	0.29	0.2	0.277
CKR_{ACT}	0.19	0.241	0.223	0.33	0.28	0.21	0.19	0.13	0.185
CKR_{TOTAL}*	0.47	0.603	0.557	0.83	0.71	0.52	0.48	0.33	0.462

Symbol (units)		**Adapted values**						
	11733	**11736**	**11813**	**11823**	**11845**	**11862**	**11873**	**11881**
T_c (h)	23	23	23	23	50	23	23	20
T_{G0} (h)	50	96	96	50	50	96	96	96
T_N (h)	20	20	20	20	10	10	20	10
T_A (h)	6	6	6	6	6	6	6	6
R_A (h^{-1})	0.001	0.001	0.001	0.001	0.0005	0.001	0.001	0.001
R_{ADiff} (h^{-1})	0.03	0.018	0.018	0.004	0.06	0.01	0.01	0.005
R_{NDiff} (h^{-1})	0.001	0.001	0.001	0.001	0.001	0.01	0.001	0.001
P_{G0toG1}	0.01	0.01	0.01	0.01	0.1	0.01	0.01	0.01
N_{LIMP}	7	7	7	7	7	7	7	7
P_{sym}	0.49	0.46	0.48	0.46	0.5	0.45	0.47	0.48
P_{sleep}	0.28	0.28	0.28	0.28	0.28	0.28	0.28	0.28
CKR_{VCR}	0.37	0.19	0.28	0.35	0.19	0.28	0.28	0.41
CKR_{ACT}	0.25	0.13	0.19	0.23	0.13	0.19	0.19	0.27
CKR_{TOTAL}*	0.62	0.32	0.47	0.58	0.32	0.47	0.47	0.68

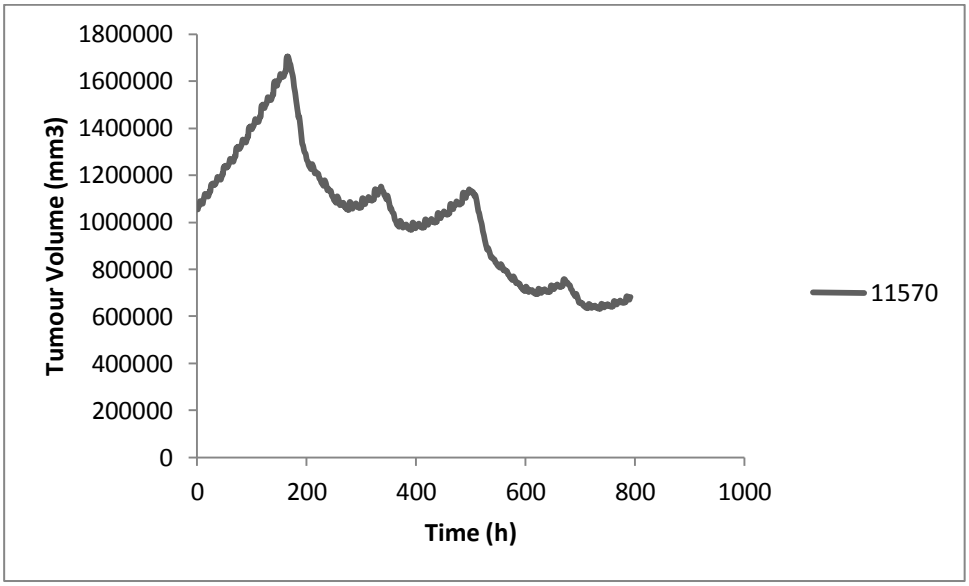
TABLE N.VIII: Initial and final virtual tumours characteristics and tumour volume reduction defined by the model parameters values presented in Table VII.

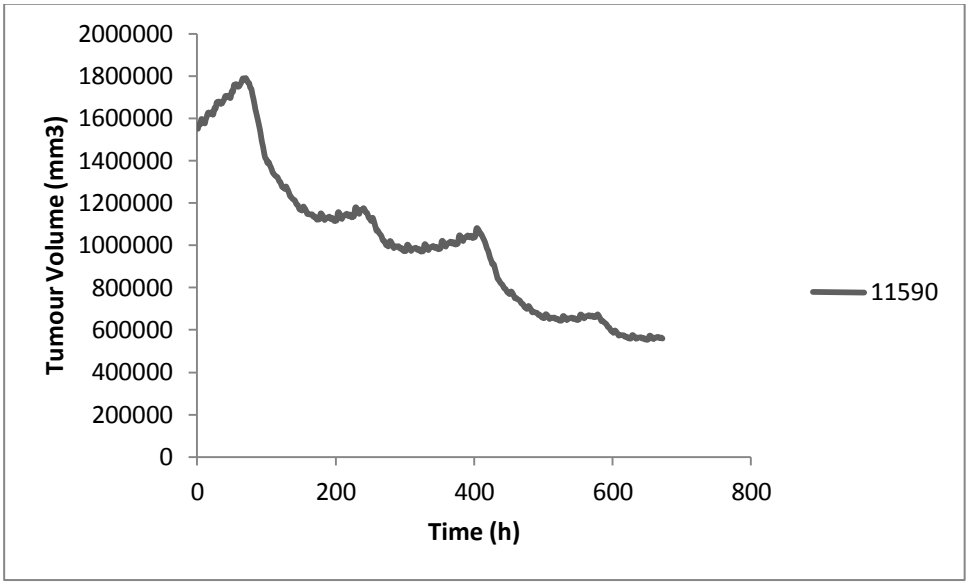
Case	11351	11570	11590	11627	11628	11639	11803	11537	11714
Volume Doubling Time	11.8	10.5	14.96	11.69	28.52	40.38	22.08	40.38	13.23
Initial percentage of proliferating cells (Growth Fraction) (%)	33.98	27.35	27.23	26.47	17.5	14.49	12.67	11.05	27.21
Initial percentage of dormant cells (%)	40.54	39.41	32.34	32.98	23.96	20.24	17	15.44	34.5
Initial percentage of differentiated cells (%)	13.88	24.2	31.21	31.27	50.95	59.34	65.61	69.03	28.44
Initial percentage of dead cells (%)	11.6	9.04	9.23	9.43	9.28	5.94	4.72	4.48	9.85
Tumour volume reduction percentage (%)	-9.78	35.66	63.84	79.82	80.17	71.63	38.86	32.06	2.58
Final percentage of proliferating cells (Growth Fraction) (%)	32.68	24.79	24.55	17.38	17.1	9.3	9.12	9.32	26.99
Final percentage of dormant cells (%)	42.12	42.47	34.14	43.84	36.51	15.68	12.63	13.07	34.64
Final percentage of differentiated cells (%)	13.22	23.09	31.35	20.82	32.69	69.73	74.31	73.54	28.54
Final percentage of dead cells (%)	11.98	9.65	9.97	17.96	13.7	5.3	3.94	4.08	9.83

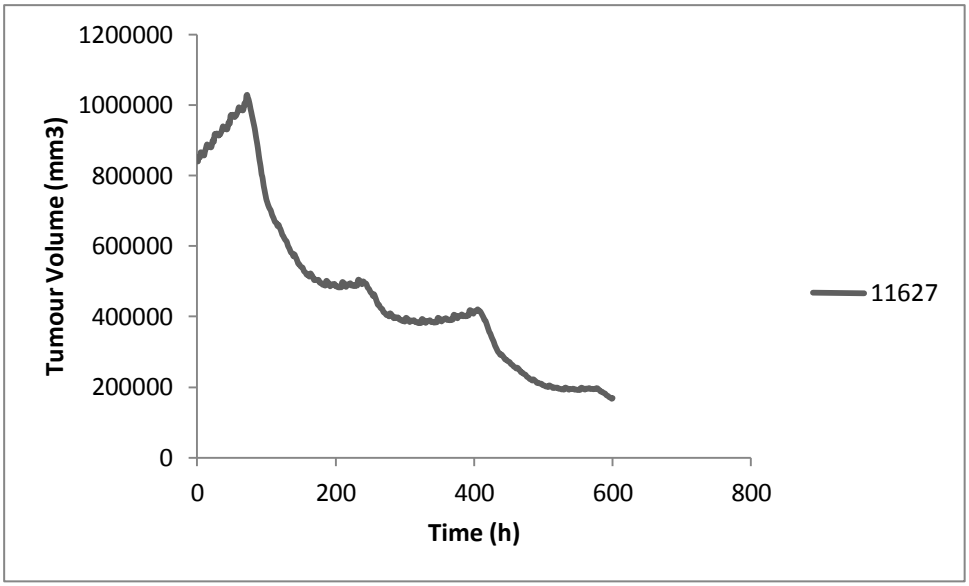
Cont. ↓

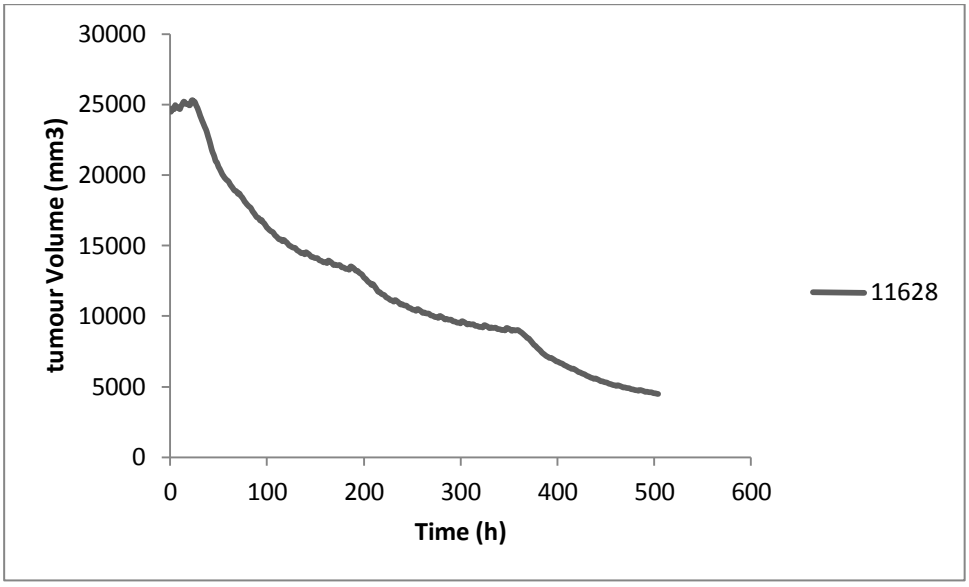
Case	11733	11736	11813	11823	11845	11862	11873	11881
Volume doubling time	13.13	22.08	15.25	21.86	25.97	11.69	18.03	12.44
Initial percentage of proliferating cells (Growth Fraction) (%)	36.02	25.87	26.92	17.22	61.1	30.93	21.76	16.1
Initial percentage of dormant cells (%)	26.35	34.7	34.76	12.97	22.21	38.52	28.62	23.17
Initial percentage of differentiated cells (%)	23.4	28.9	27.96	62.12	9.12	19.52	41.42	54.45
Initial percentage of dead cells (%)	14.23	10.53	10.36	7.69	7.56	11.03	8.2	6.28
Tumour volume reduction percentage (%)	72.81	25.01	24.56	70.49	34.06	70.21	41.75	73.61
Final percentage of proliferating cells (Growth Fraction) (%)	33.93	24.52	43.53	10.31	59.39	30.07	20.79	9.99
Final percentage of dormant cells (%)	27.48	35.72	36.66	8.66	22.77	39.38	28.65	16.86
Final percentage of differentiated cells (%)	24.05	29.23	28.12	74.56	8.74	19.35	42.3	67.6
Final percentage of dead cells (%)	14.53	10.53	10.66	6.46	9.11	11.2	8.26	5.55

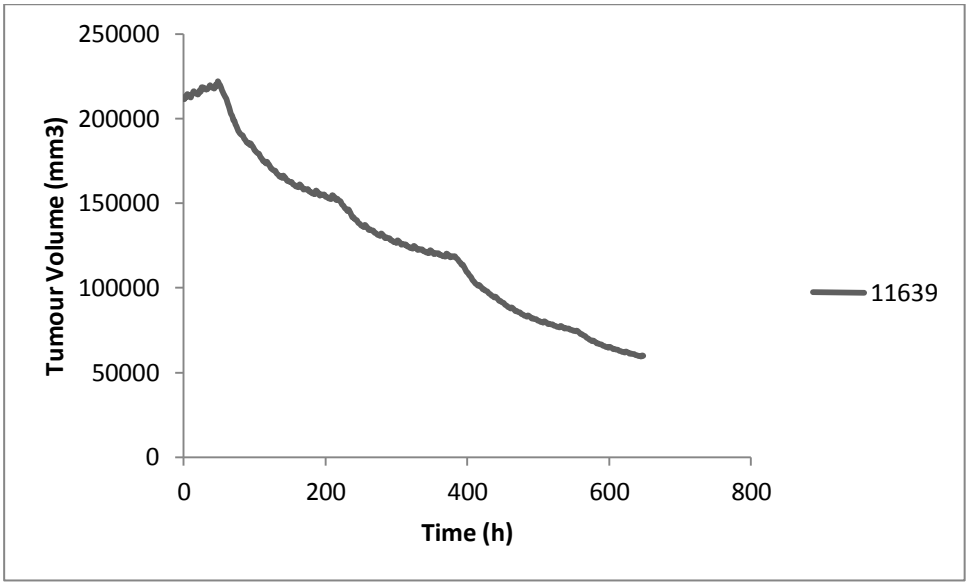
Fig. N.3 ↓ Time evolution of tumour volume for virtual tumours defined in Table N.VII. (16 panels)

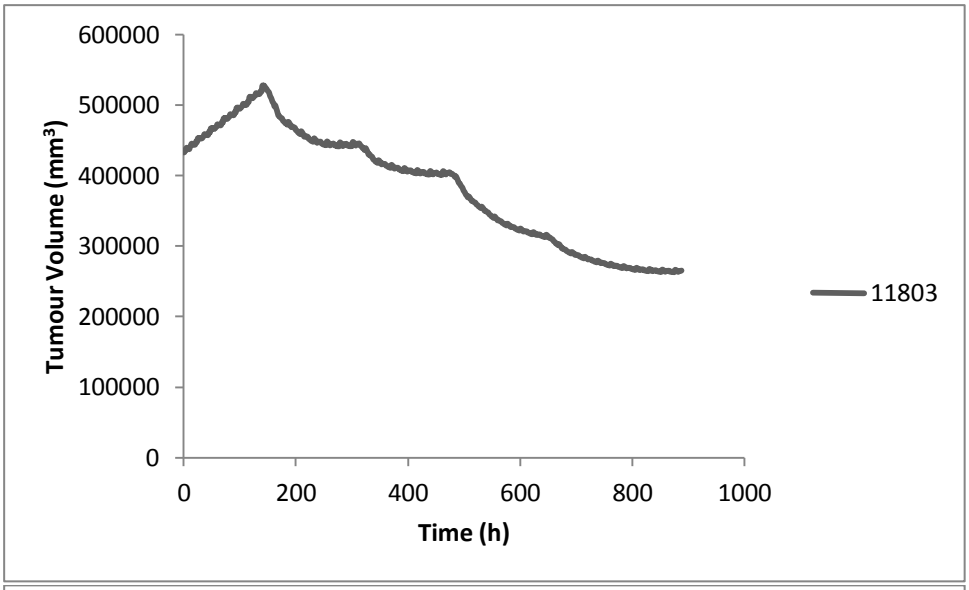


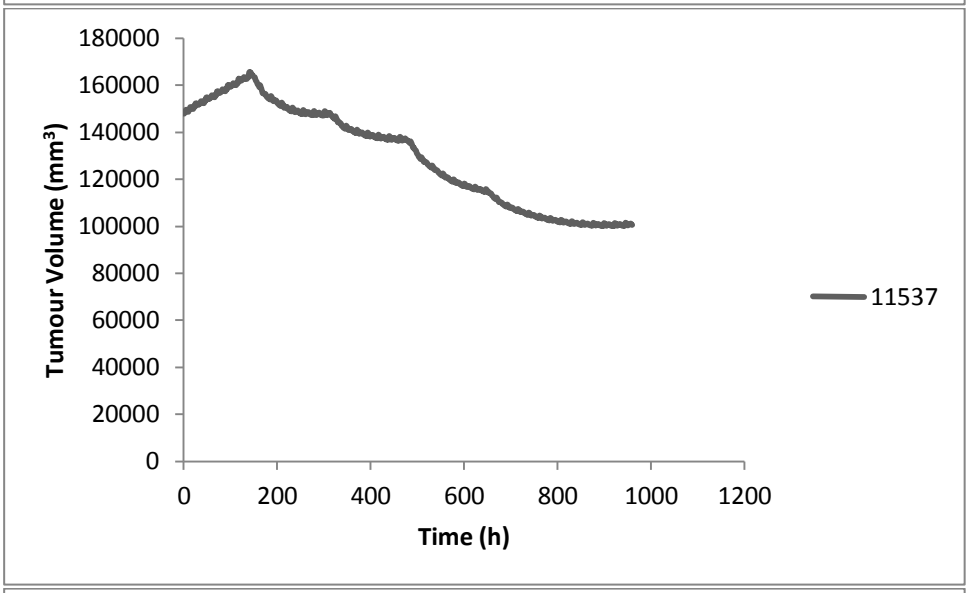


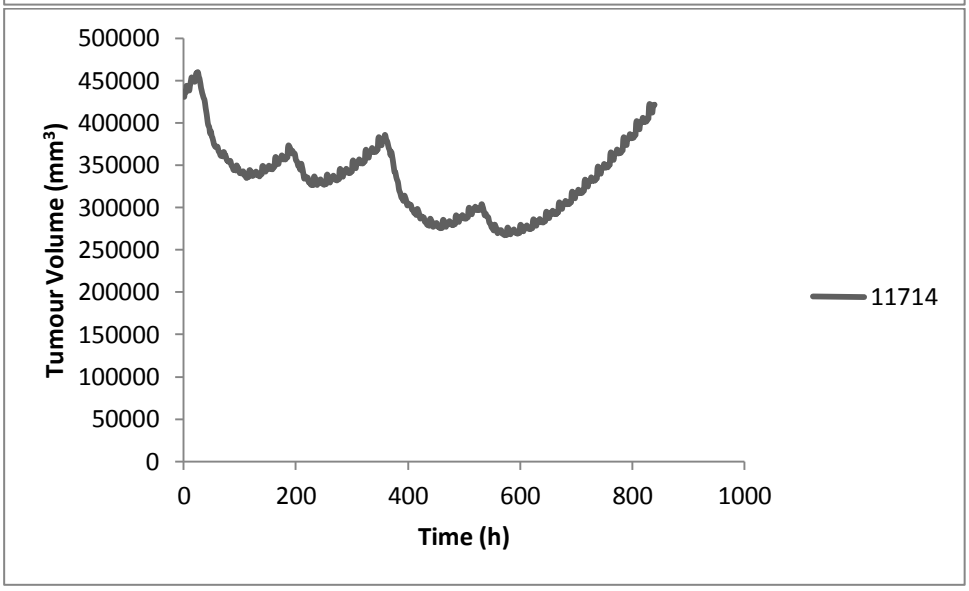

1200000
1000000
800000
600000
400000
200000
0
Tumour Volume (mm3)
0
200
400
600
800
Time (h)
11627

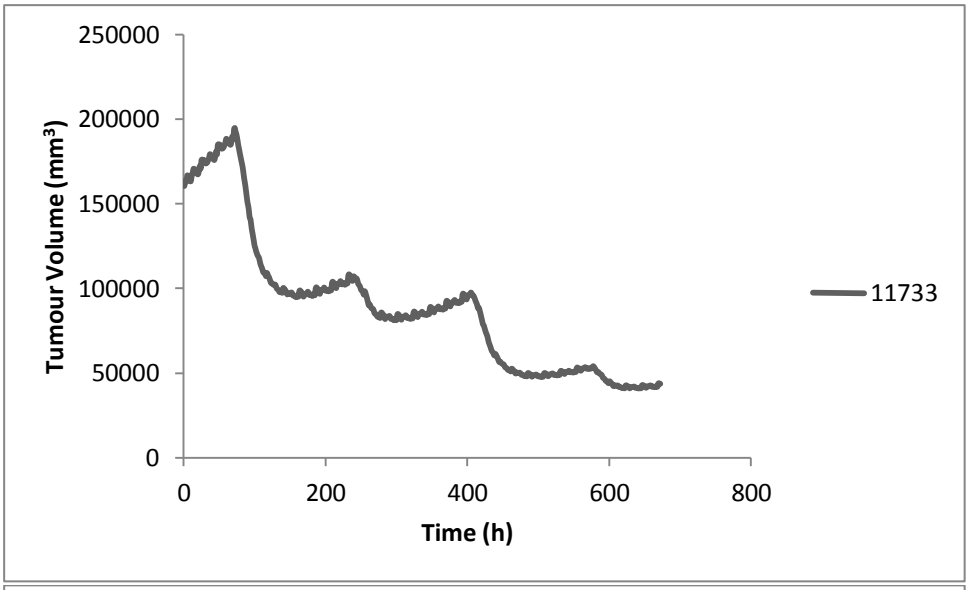

30000
25000
20000
15000
10000
5000
0
tumour Volume (mm3)
0
100
200
300
400
500
600
Time (h)
11628

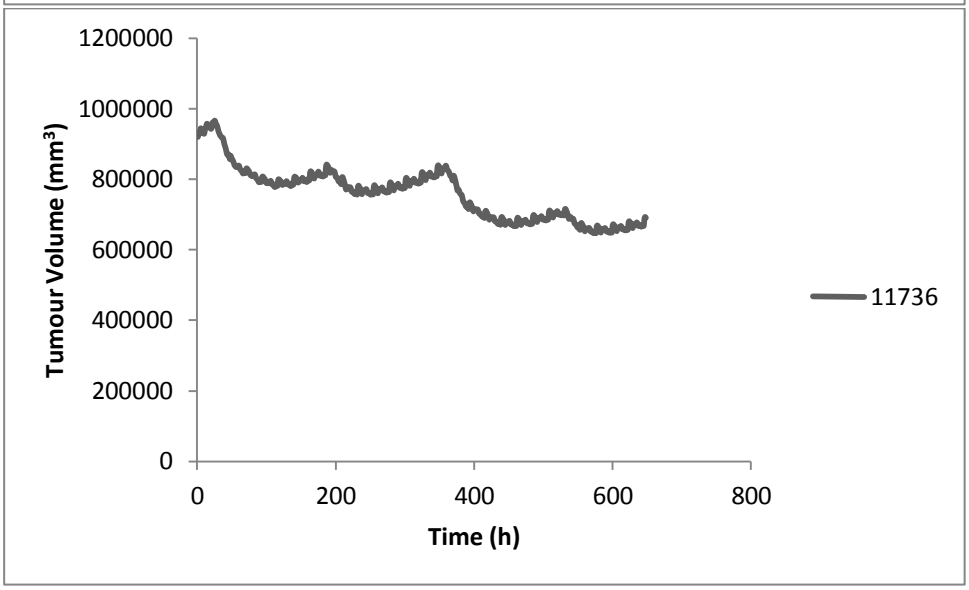

250000
200000
150000
100000
50000
0
Tumour Volume (mm3)
0
200
400
600
800
Time (h)
11639

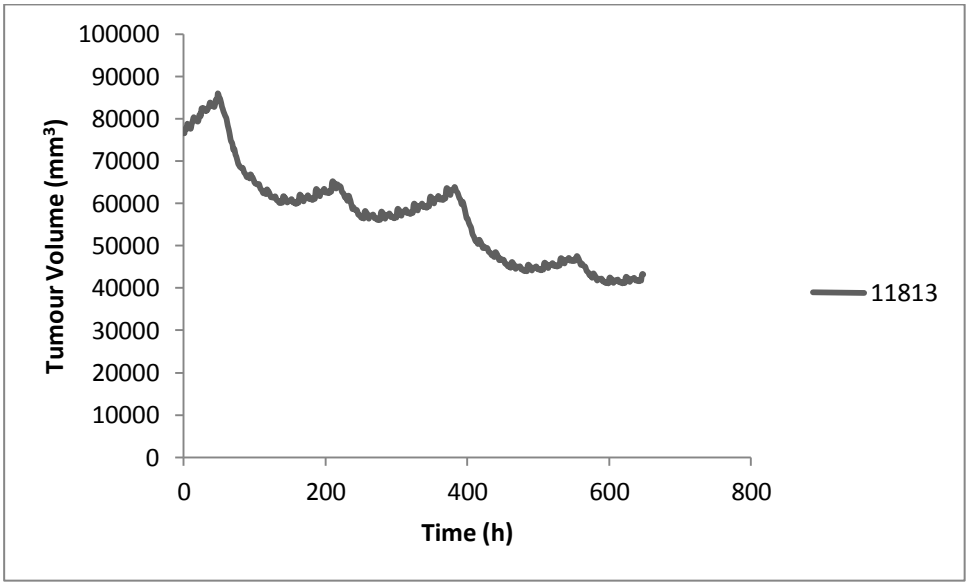

600000
500000
400000
300000
200000
100000
0
Tumour Volume (mm³)
0
200
400
600
800
1000
Time (h)
11803

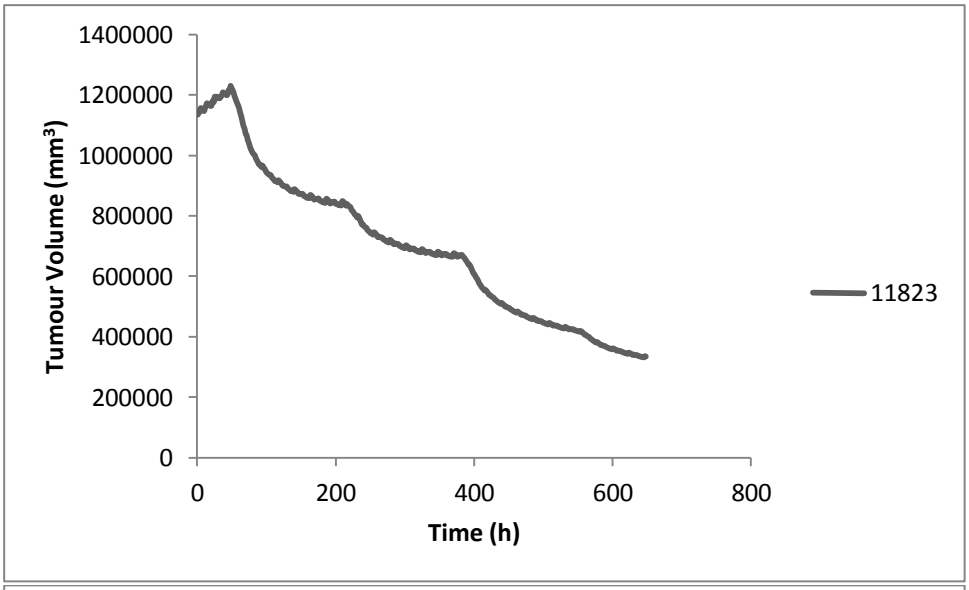

180000
160000
140000
120000
100000
80000
60000
40000
20000
0
Tumour Volume (mm³)
0
200
400
600
800
1000
1200
Time (h)
11537

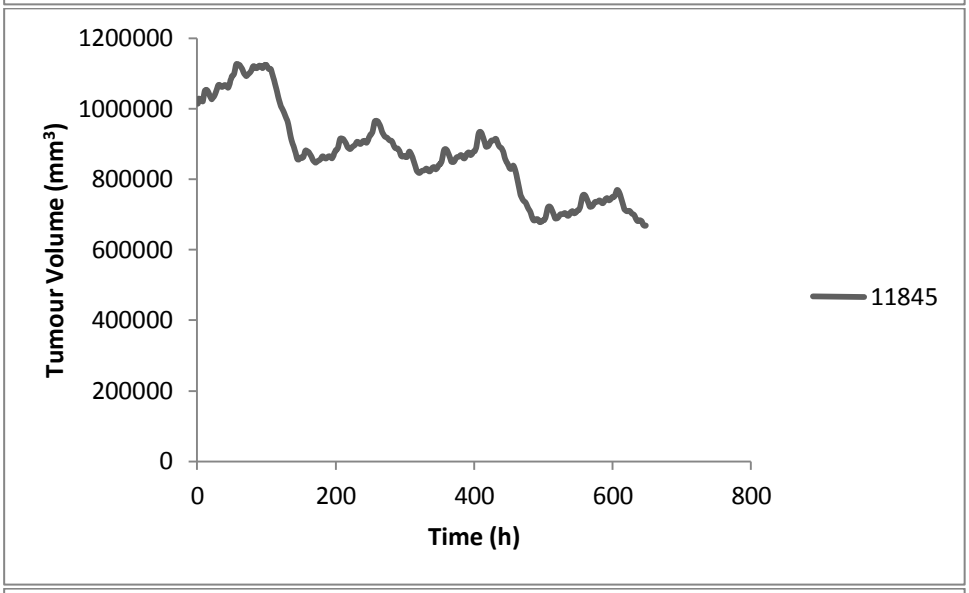

500000
450000
400000
350000
300000
250000
200000
150000
100000
50000
0
Tumour Volume (mm³)
0
200
400
600
800
1000
Time (h)
11714

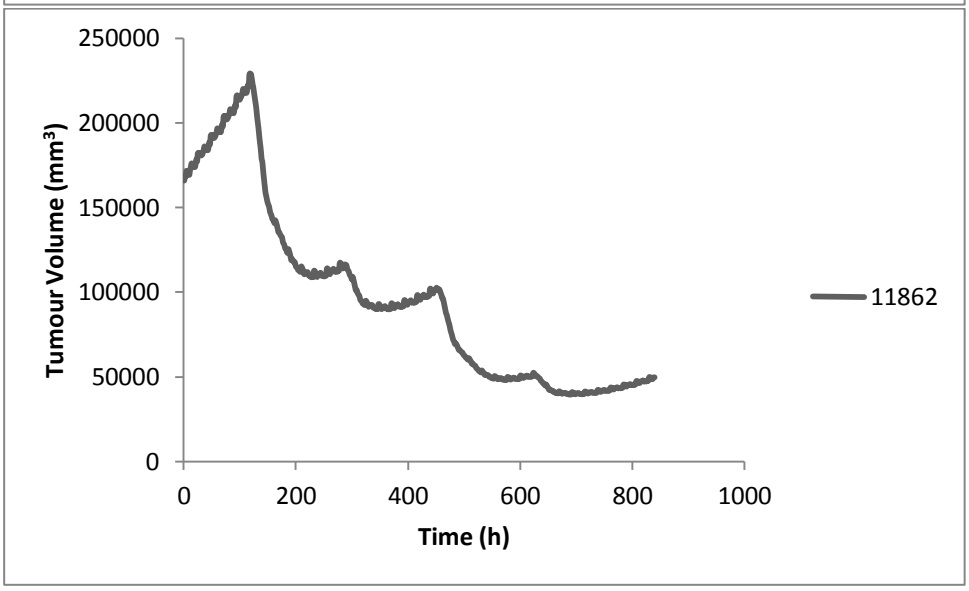

250000
200000
150000
100000
50000
0
Tumour Volume (mm³)
0
200
400
600
800
Time (h)
11733


1200000
1000000
800000
600000
400000
200000
0
Tumour Volume (mm³)
0
200
400
600
800
Time (h)
11736


100000
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
Tumour Volume (mm³)
0
200
400
600
800
Time (h)
11813


1400000
1200000
1000000
800000
600000
400000
200000
0
Tumour Volume (mm³)
0
200
400
600
800
Time (h)
11823


1200000
1000000
800000
600000
400000
200000
0
Tumour Volume (mm³)
0
200
400
600
800
Time (h)
11845


250000
200000
150000
100000
50000
0
Tumour Volume (mm³)
0
200
400
600
800
1000
Time (h)
11862

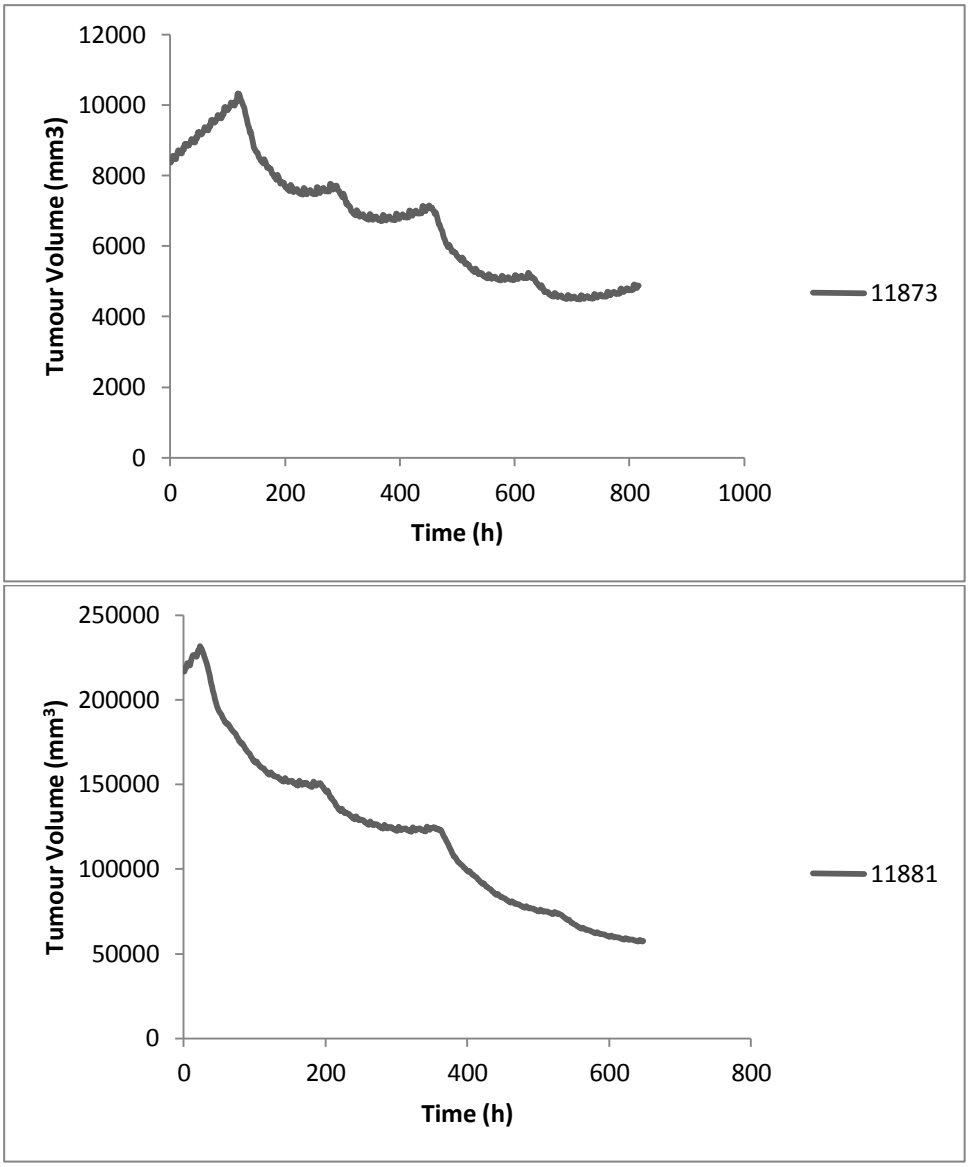


Fig. N.3: Time evolution of tumour volume for 17 virtual tumours defined in Table N.VII.

The fitting of the selected 17 nephroblastoma cases to the clinical data demonstrates the basic philosophy of the model adaptation. This is based on the combined use of the available multiscale clinical and literature data. The availability of multiscale medical data imposes constraints on the possible model parameter values. A possible set of model parameters has been assigned to each case leading to a virtual tumour with resulting characteristics in accordance with the available medical data.

As the available information regarding a particular tumour characteristics becomes richer and richer, a narrowing of the window of possible solutions is expected. Additionally, efforts to achieve an adequate ‘‘tuning’’ of the simulation results to reality may provide valuable hints concerning the tumour characteristics for which parameter value estimates may be missing. Currently efforts to exploit the information provided by microRNA differential expression profiles is under way. In this way estimates of the internal histological constitution of nephroblastoma tumours may be produced before the start of chemotherapy. This is expected to considerably improve the reliability of the N Oncosimulator.

N.4 REFERENCES

[N.1] G Stamatakos, D Dionysiou, A Lunzer, R Belleman, E Kolokotroni, E Georgiadi, M Erdt, J Pukacki, S Rueping, S Giatili, A d`Onofrio, S Sfakianakis, K Marias, C Desmedt, M Tsiknakis, and N Graf "The Technologically Integrated Oncosimulator: Combining Multiscale Cancer Modelling with Information Technology in the In Silico Oncology Context" IEEE J Biomed Health Inform 2013

[N.2] E.C.Georgiadi, D.D.Dionysiou, N.Graf, G.Stamatakos, "Towards In Silico Oncology: Adapting a Four Dimensional Nephroblastoma Treatment Model to a Clinical Trial Case Based on Multi-Method Sensitivity Analysis." Computers in Biology and Medicine, in press, doi: 10.1016/j.compbiomed.2012.08.008. Epub 2012 Oct 10. vol.42 (11) pp. 1064-1078 2012

[N.3] G.S.Stamatakos, E.Ch.Georgiadi, N.Graf, E.A.Kolokotroni, and D.D.Dionysiou, "Exploiting Clinical Trial Data Drastically Narrows the Window of Possible Solutions to the Problem of Clinical Adaptation of a Multiscale Cancer Model", PLOS ONE 6(3), e17594, 2011

[N.4] N. Graf, A. Hoppe , E. Georgiadi, R. Belleman, C. Desmedt, D. Dionysiou, M. Erdt , J. Jacques, E. Kolokotroni, A. Lunzer, M. Tsiknakis, G. Stamatakos "In Silico Oncology" for Clinical Decision Making in the Context of Nephroblastoma. [Die Bedeutung von ‚in silico Onkologie` zur klinischen Entscheidungsfindung am Beispiel des Nephroblastoms] Klin. Paediatr. (Klinische Paediatrie) 221, 141-149, 2009

[N.5] E.A. Kolokotroni, D.D. Dionysiou, N.K. Uzunoglu, G.S. Stamatakos, "Studying the growth kinetics of untreated clinical tumours by using an advanced discrete simulation model", Mathematical and Computer Modelling, 54 1989-2006 2011

[N.6] Berrebi D, Leclerc J, Schleiermacher G, Zaccaria I, Boccon-Gibod L, et al. (2008) High cyclin-E staining index in blastemal, stromal or epithelial cells is correlated with tumour aggressiveness in patients with nephroblastoma. PLoS ONE 3(5): e2216. doi:10.1371/journal.pone.0002216.

[N.7] Tan TY, Amor DJ (2006) Tumour surveillance in Beckwith-Wiedermann syndrome and hemihyperplasia: A critical review of the evidence and suggested guidelines for local practice. Journal of Paedeatrics and Child Health 42: 486–490.

[N.8] Shackney SE, McCormack GW, Cuchural GJ (1978) Growth rate patterns of solid tumours and their relation to responsiveness to therapy. Ann Intern Med 89: 107–121.

[N.9] Craft AW (1999) Growth rate of Wilms' tumour. The Lancet 354(9184): 1127.

[N.10] Zoubek A, Slavc I, Mann G, Trittenwein G, Gadner H (1999) Natural course of a Wilms' tumour. Lancet 354: 344.

[N.11] Carre´ A, Frantz CN, Weksberg R, Nicholson L, Ciarlo L, et al. (2005) Wilms tumour in an 11-year old with hemihyperplasia. Am J Med Genetics 139A:165–166

[N.12] Steel GG (2002) Basic Clinical Radiobiology. London: Arnold. pp 9–10.

B. Report on the clinical adaptation and optimization of the breast cancer Oncosimulator models. *In silico* studies.

The objective of the breast cancer branch of the Oncosimulator is to simulate the response of clinical breast tumours to specific treatment schemes and/or schedules in the patient individualized context. To this end, a continuum approach describing vascular tumour growth under angiogenic signaling has been developed based on relevant literature (Poleszczuk *et al.*, 2011; Hahnfeldt *et al*., 1999) and bevacizumab pharmacokinetic properties have been incorporated (Argyri *et al.*, 2012). Thus, the mono-therapy effect for the specific anti-angiogenic inhibitor, which probably is the most popular representative of the wider family of antiangiogenic agents in clinical practice, has been simulated and the results produced by the continuum approach have been thoroughly studied. Additionally, a discrete entity – discrete event breast cancer model addressing the chemotherapeutic effect has been developed (see Chapter 3 in deliverable D12.3) and vinorelbine pharmacokinetic properties have been modelled. Finally, a coupling approach between the dynamical system describing vascular tumour growth and the cytokinetic breast cancer model that stems from previous work of the *In Silico* Oncology Group (ISOG) (Stamatakos *et al.,* 2010) is in progress in order to account for the effect of concomitant administration of anti-angiogenic treatment and cytotoxic agents.

The main focus of the chapter is twofold: first to present an exploratory parametric analysis that has been conducted in order to gain insight into the free growth dynamics of a vascularized tumour. Additionally, to present the local fitting of both the free growth and the treatment module of the vascular tumour growth model to actual experimental data concerning the *in vivo* study of solid tumours that has been performed and to outline the ongoing process of global fitting so as to acquire an initial confirmation that the underlying mathematical model (Hahnfeldt *et al.*, 1999; Poleszczuk *et al*., 2011; Argyri *et al.*, 2012) is indeed tailored to the nature of the problem under study.

B.1 Basic features and internal information flow of the vascular tumour growth model

The scientific problem to be addressed consists not only of the description of tumour development but also of the description of the key-process called angiogenesis, the anti-angiogenic treatment effect and last but certainly not least, the elucidation of the interdependencies among the aforementioned three processes. To this end, the mathematical model describing the problem under study has been based on the approach of (Poleszczuk *et al*., 2011) which stems from previous work of (Hahnfeldt *et al*., 1999).

All major biological phenomena of cancer cell population dynamics are incorporated into the model i.e. cancer cell proliferation, cancer cell apoptosis, post-vascular dormancy (state where tumour growth ceases due to the balance achieved between pro-angiogenic and antiangiogenic factors), endothelial cell death, spontaneous loss of functional vasculature, excretion of endogenous proangiogenic factors (such as vascular endothelial growth factor, fibroblast growth factors, platelet-derived growth factor, angiopoietin-1 etc.), excretion of endogenous anti-angiogenic factors (angiostatin, endostatin, angiopoietin-2 etc.) and anti-angiogenic treatment – induced endothelial cell death as well as the resulting cancer cell death.

The implicit assumptions on which the basic framework of the model is based are that the tumour is a three dimensional spheroid, the diffusion process is in a quasi-stationary state i.e. the tumour growth rate as well as the rate of change of drug concentration are relatively small compared to the

rate of distribution of angiogenesis stimulators and the concentration of the stimulator is a radially symmetric function.

The model of (Poleszczuk *et al*., 2011) makes use of the concept of a variable carrying capacity i.e. the maximal tumour volume that can be supported by the given vasculature which was originally introduced in (Hahnfeldt *et al*., 1999). The dynamical system described in (Poleszczuk *et al*., 2011) consists of a pair of ordinary differential equations (ODEs) which reflect the interplay between tumour volume (*V*) and carrying capacity (*K*).

The model of vascular tumour growth under anti-angiogenic treatment consists of three – components: a tumour compartment monitoring the rate of change of the tumour volume *V* under the assumption of Gompertzian growth, a vascular compartment keeping track of the temporal evolution of the carrying capacity of the cancer cell population and finally an anti-angiogenic treatment compartment monitoring the time-course of bevacizumab concentration in plasma based on a two-compartmental pharmacokinetic model (Bertrand and Mentré, 2008). The authors proceeded with the implementation of the model in MATLAB. The details concerning the m-files involved in the implementation of the model are listed and explained below.

DrugConcentration.m: This function implements the two-compartmental pharmacokinetic model for administration via intravenous infusion (as it is the case for bevacizumab in clinical practice). It calculates the concentration of the anti-angiogenic agent at each time point taking into account all previous infusions.

VascularTumourGrowth.m: This function computes the derivatives involved in the ODEs describing tumour growth rate and carrying capacity rate of change (Fig. B.1).

VascularTumourGrowth_main.m: Given the initial values (V_0, K_0) and a time interval, the specific script file resolves the problem with the solver ode45 which implements a Runge Kutta method with a variable time step for efficient computation. It also plots the variables *V* and *K* (representing tumour volume and tumour capacity, respectively) as functions of time in common axial system as well as the phase – plane of the system.

It is worth noting that the experiments of free growth, constant treatment and intermittent treatment have been performed on a desktop computer with an AMD Phenom(tm) II X6 1055T Processor 2.80 GHz, 8.00GB RAM in Windows 7 with 64 – bit operating system and that the code execution time is of the order of a few seconds.

Fig. B.1 demonstrates the information flow inside the vascular tumour growth under anti-angiogenic treatment model.

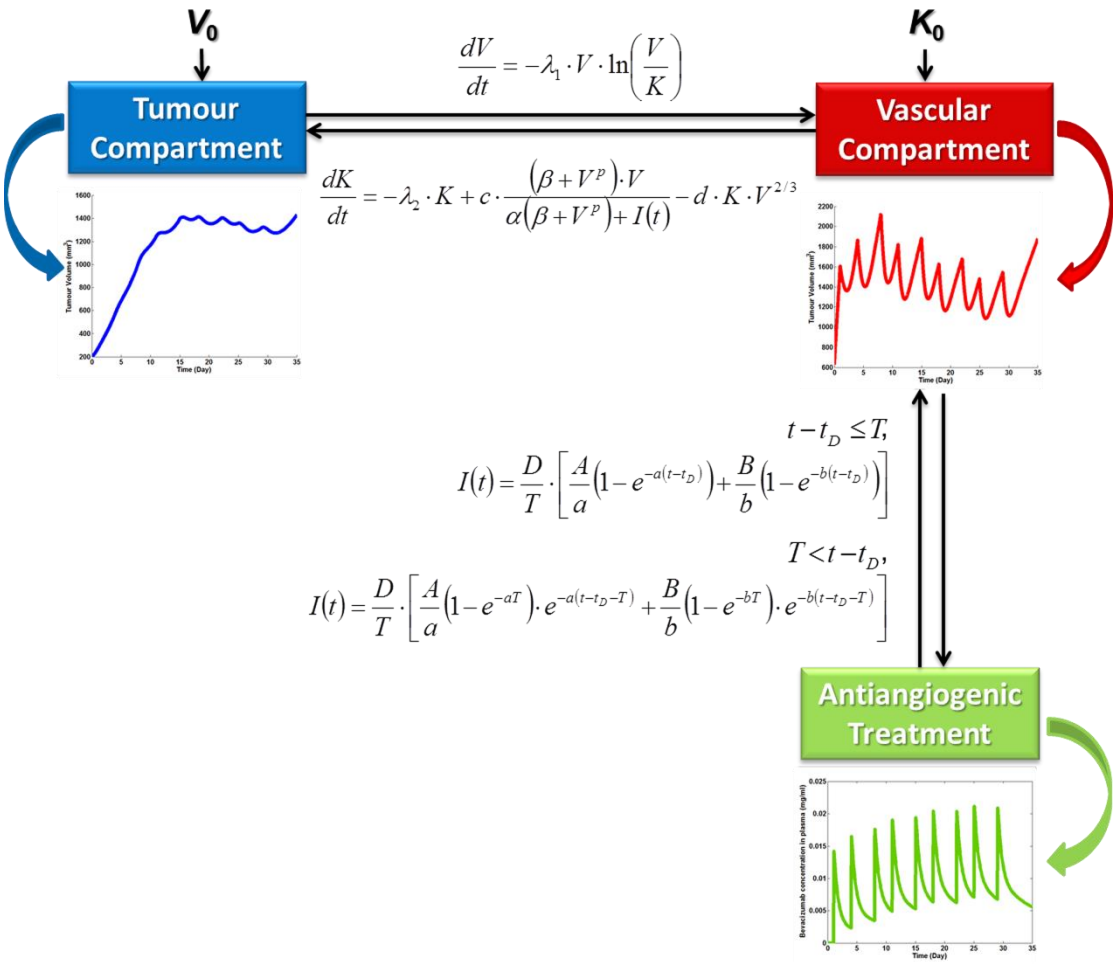


Fig. B.1: The information flow inside the vascular tumour growth under anti-angiogenic treatment model

The parameters and the variables involved in the equations of the vascular tumour growth model are explained in Table B.1 while the parameters and the variables of the two compartmental pharmacokinetic model are explained in Table B.2.

TABLE B.1 Variables and parameters related to the vascular tumour growth model.

	Mathematical Symbol	Units	Description
Independent variable	t	days	time
Dependent Variable	V	mm^3	tumour volume
Dependent Variable	K	mm^3	tumour capacity
Parameter	λ_1	day^{-1}	Gompertzian growth constant
Parameter	λ_2	day^{-1}	proportionality constant related to the term reflecting the spontaneous loss of functional vasculature
Parameter	c	$mg/(ml \cdot day \cdot mm^{3p})$	proportionality constant related to the term reflecting endogenous stimulation of the tumour upon the vasculature
Parameter	d	$1/(day \cdot mm^2)$	proportionality constant related to the term reflecting endogenous inhibition of tumour vasculature
Parameter	α	$mg/(mm^{3p} \cdot ml)$	constant related to the total amount of stimulators inside the tumour
Parameter	β	mm^{3p}	constant related to the total amount of stimulators inside the tumour
Parameter	p	≥0	Hill coefficient
Parameter	I	mg/ml	drug concentration in plasma

TABLE B.2 Variables and parameters related to the two–compartmental pharmacokinetic model that computes bevacizumab concentration in plasma for a given time-point (Bertrand and Mentre, 2008).

	Mathematical Symbols	Units	Description
Independent variable	t	day	time-point
Dependent variable	I	mg/ml	bevacizumab concentration in plasma
Parameter	n	-	number of infusions to be administered
Parameter	D	mg	dose
Parameter	T	day	infusion duration
Parameter	V_p	ml	volume of central compartment
Parameter	k_{12}	day^{-1}	transfer constant from central to peripheral compartment
Parameter	k_{21}	day^{-1}	transfer constant from peripheral to central compartment
Parameter	k_e	day^{-1}	rate of elimination
Parameter	-	mg/kg	dosage
Parameter	-	kg	patient's weight
Parameter	t_i	day	administration time-points of anti-angiogenic treatment

B.2 Results of an exploratory set of parametric simulations for the case of an untreated tumour

A parameter analysis has been conducted in order to investigate the impact of the input factors on the output of the model as well as to identify the model parameters exerting the greatest influence on the model predictions. To this end, firstly a *varying one parameter at a time* approach has been selected and subsequently the combined effect of all combinations of input parameters has been studied so as to provide insight into the dynamic behaviour of an untreated solid tumour.

In the free growth module, the vascular tumour growth model reduces to the following form:

$$\frac{dV}{dt} = -\lambda_1 \cdot V \cdot \ln\left(\frac{V}{K}\right) \quad \text{(B.1)}$$

$$\frac{dK}{dt} = -\lambda_2 \cdot K + c \cdot V - d \cdot K \cdot V^{2/3} \quad \text{(B.2)}$$

Taking into account the fitting of the Hahnfeldt *et al*. model to the experimental data where λ_2 has been calculated as equal to 0, the term reflecting spontaneous loss of functional vasculature has been omitted. Consequently, the input parameters to be varied are λ_1, *c* and *d*.

The impact of the input parameters is measured with respect to their effect to the value of the plateau reached by the tumour (in mm^3) as well as to the time-point of its attainment (in days). To this end, the formulated two assumptions are noted:

- The tumour is assumed to have reached the plateau when the whole number part of the value of tumour volume coincides with the respective one of carrying capacity for at least ten consecutive simulation time-steps.
- The time-point that corresponds to the first of the aforementioned time-steps is considered as the exact time of plateau and the repeated value as the value of the plateau.

The aforementioned end-points have been chosen so as to reflect the dynamics of the untreated disease. Were the parameter investigation conducted in the context of the therapy module, different end – points should also be selected and specifically end-points interpretable as efficacy measures of the administered anti-angiogenic treatment, such as tumour growth inhibition or tumour volume reduction.

As an initial step, the perturbation of one parameter at a time by a uniform 20 percent (%) either side of the reference value (Hahnfeldt *et al*., 1999) has been performed.

Parameter λ_1: A proportionality constant with units of 1/day involved in the equation describing the rate of change of tumour volume *V*. λ_1 is the only parameter directly involved in the mathematical description of vascular tumour growth rate and as such it implicitly reflects the doubling time of the simulated tumour.

All parameters have been kept constant while λ_1 has been perturbed by a uniform 20 percent either side of the base value (blue line) which has been set equal to 0.192 1/day (Hahnfeldt *et al.*, 1999). Values 0.192, 0.2304 and 0.1536 1/day have been assigned to λ_1 in the context of the parameter analysis.

Taking equation (B.1) into account, it is expected that for untreated disease higher values of λ_1 would result in a more aggressive tumour. Indeed, as demonstrated in Fig. B.2, increasing the value of λ_1 leads to a faster growing tumour, while decreasing the value of λ_1 has the opposite effect, namely it leads to a less aggressive tumour. From the viewpoint of the selected endpoints, a closer inspection of Fig. B.2 reveals the impact of parameter λ_1 to the actual time-point that the tumour attains the plateau. On the contrary, the specific parameter does not seem to affect the value of the plateau. The measured effect is presented in Table B.3.

Parameter *c*: A proportionality constant involved in the equation describing the rate of change of the carrying capacity and specifically in the term representing the stimulatory capacity of the tumour upon the inducible vasculature, with units of 1/day. All parameters have been kept constant, while *c* has been perturbed by a uniform 20 percent either side of the base value (blue line), which has been set equal to 5.85 1/day (Hahnfeldt *et al*., 1999). Values 5.85, 7.02 and 4.68 1/day have been thus assigned to *c*. According to equation (B.2) an increased value of parameter *c* would result in a higher angiogenic stimulatory capacity of the tumour and thus to a faster growing tumour. Similarly, decreasing *c* would result in a more slowly growing tumour, a conclusion that is also corroborated in a graphical way (Fig. B.2). The great impact of perturbing *c* in the actual value of the plateau attained by the tumour is presented in Table B.3.

Parameter *d*: A proportionality constant involved in the equation describing the rate of change of the carrying capacity and specifically in the term representing the endogenous inhibition of previously generated vasculature with units of $day^{-1} \cdot mm^{-2}$. All parameters have been kept constant, while *d* has been perturbed by a uniform 20 percent either side of the base value (blue line), which has been set equal to 0.00873 $day^{-1} \cdot mm^{-2}$ (Hahnfeldt *et al*., 1999). Values 0.00873, 0.010476 and 0.0069841/day have been thus assigned to *d*. According to equation (B.2), high values of parameter *d* lead to tumours with high endogenous inhibition of previously generated vasculature and faster growing tumours. In a similar way, lower values of *d* correspond to tumours with small concentrations of angiogenesis inhibitors and thus, to slower growing tumours (Fig. B.2). The significant impact of perturbing *d* in the actual value of the plateau attained by the tumour is presented in Table B.3.

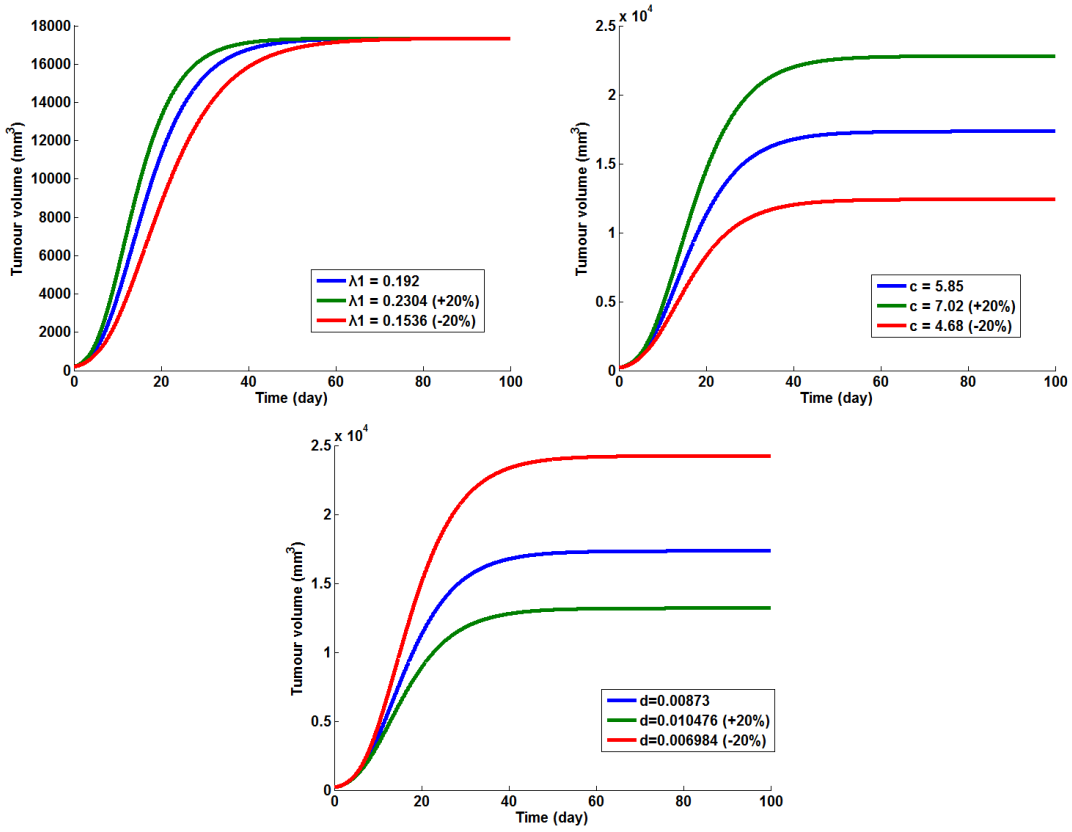


Figure B.2: The effect of perturbating parameters λ_1, *c* and *d* respectively on the volume of the untreated tumour when all other parameters are kept constant at the reference values adopted in (Hahnfeldt *et al.*, 1999).

TABLE B.3 Effect of varying the input parameter value on the value of the tumour plateau as well as on the time-point of its attainment

Parameter values			Units	Time of plateau (days)	Value of plateau (mm^3)	Variation of the plateau value (%)
λ_1	reference value	0.192	day^{-1}	90	17346	n/a
	+20%	0.2304		76	17346	0
	-20%	0.1536		112	17346	0
c	reference value	5.85	day^{-1}	90	17346	n/a
	+20%	7.02		92	22802	+31.45%
	-20%	4.68		90	12412	**-28.44%**
d	reference value	0.00873	$day^{-1} \cdot mm^{-2}$	90	17346	n/a
	+20%	0.010476		84	13195	-23.93%
	-20%	0.006984		93	24242	**+39.76%**

As it becomes obvious from examining Fig. B.2 and Table B.3, *c* and *d*, related to the total amount of angiogenic stimulators and inhibitors respectively, are the main modulators of the value of the plateau attained by the tumour, while λ_1 is the main modulator of the time-point that the tumour reaches the plateau.

In fact, two extra sets of simulations have been conducted, aiming at determining in a graphical way the shortest path to a more aggressive and to a less aggressive tumour respectively, varying only one parameter at a time. Hence, we proceed with the demonstration of the effect of perturbing λ_1 (green curve) and *c* (red curve) by a uniform +20% and *d* (light blue curve) by a uniform -20% of the respective reference values (Fig. B.3) in common axial system so as to construct a more aggressive tumour. Following this, the effect of perturbing λ_1 (green curve) and *c* (red curve) by a uniform -20% and *d* (light blue curve) by a uniform +20% of the respective reference values (Fig. B.3) in order to construct a less aggressive tumour is presented.

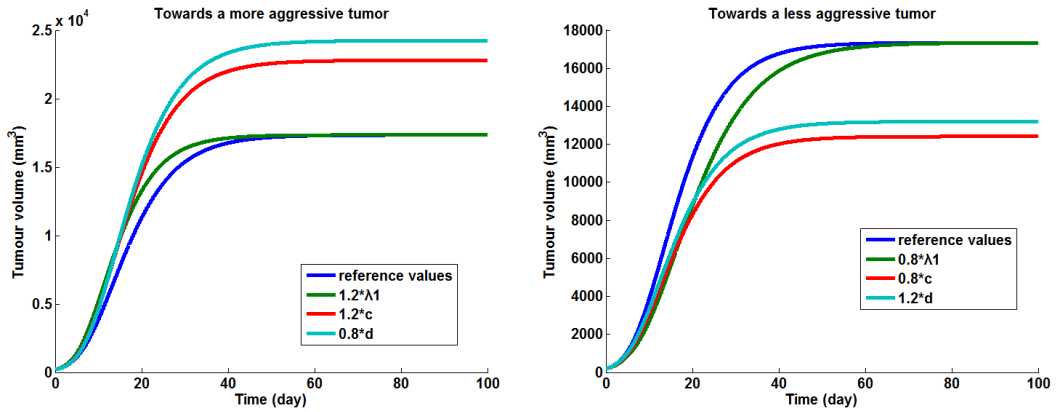


Figure B.3: The effect of perturbing λ1, c and d by a 20% in order to construct a more aggressive tumour and a less aggressive tumour respectively.

The results that have come up through graphic representation suggest that *c*, i.e. the parameter that reflects the concentration of angiogenesis stimulators, is the most sensitive parameter to changes that create a more aggressive tumour, while *d*, i.e. the parameter that relates to angiogenic stimulatory capacity, seems to be the most sensitive parameter to changes that create a less aggressive tumour. This may be suggesting that anti-angiogenic treatment aiming at reducing the amount of angiogenic stimulators (i.e. anti-angiogenic agent bevacizumab) could be a more efficient approach than increasing the amount of angiogenic inhibitors in the tumour.

However, a closer inspection of the two comparative graphs reveals a dependence of the sensitivity on the final time- point considered. For example, if one considers a final time-point in the exponential phase of the tumour (e.g. day 15 at 'towards a less aggressive tumour' scenario), λ_1 is the most influential input parameter.

Due to the limited number of parameters involved in the free growth module of the model, it has become feasible to conduct a study of the effect of varying all existing combinations of input parameters. Extending the parametrer investigation in such way can prove very significant, especially in cases where a positive or a negative correlation between input parameters may exist. For example, a tumour secreting a large amount of angiogenic stimulators is probably characterized by a smaller doubling time, a fact suggesting that perturbing parameters λ_1 and *c* at the same time would be of potential interest. Similarly, a tumour characterized by large endogenous inhibition of tumour vasculature would be expected to have greater doubling time and thus studying the effect of perturbing both λ_1 and *d* could prove useful.

The combined effect of λ_1 and *c*, λ_1 and *d, c* and *d* as well as λ_1, *c* and *d* towards obtaining a more aggressive and a less aggressive tumour has been demonstrated (Fig. B.4) using a 20% perturbation either side of the reference value. As the results of the specific parameter investigation suggest (Fig. B.4, Table B.4), (*c*, *d*) seems to be the most influential couple, a conclusion expected as the specific dyad is comprised of the two parameters that, according to the one-at-a-time parameter analysis, exert the greatest impact on the plateau value attained by the tumour (Fig. B.3, Table B.3). A closer inspection of Table B.3 and Table B.4 also corroborates the prior conclusion that λ_1 does not affect the value of the plateau. Indeed, comparing the respective columns in Table B.3 and Table B.4, it becomes obvious that perturbing the dyads (λ_1, *c*) and (λ_1, *d*) has exactly the same effect as varying only *c* and *d* respectively, with respect to the value of plateau. However, parameter λ_1 seems to modulate the timing of plateau either by accelerating it or delaying it depending on whether one increases or decreases its value respectively. As expected, perturbing all parameters involved in the vascular tumour growth model at the same time demonstrates the most pronounced influence of all

other parameter combinations tested so far whether the aim is a more aggressive tumour or the opposite.

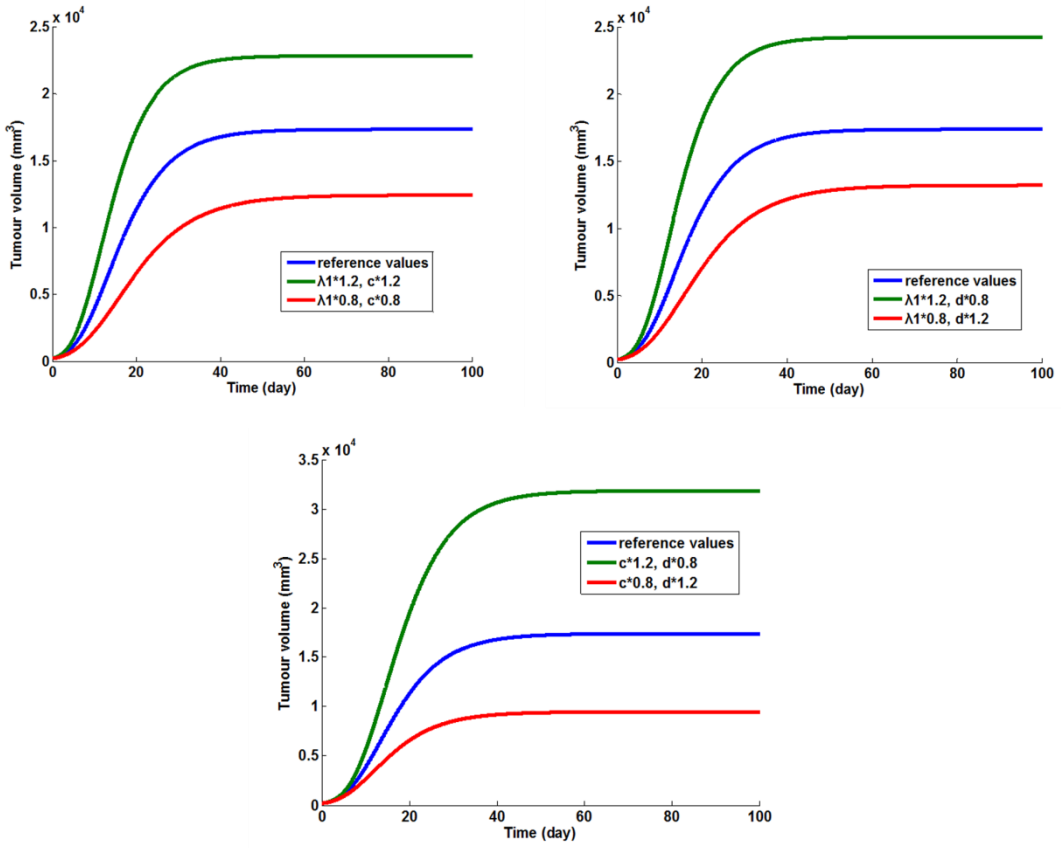


Figure B.4: The combined effect of perturbing all parameter dyads by a 20% so as to construct a more aggressive tumour (green line) and a less aggressive tumour (red line) respectively.

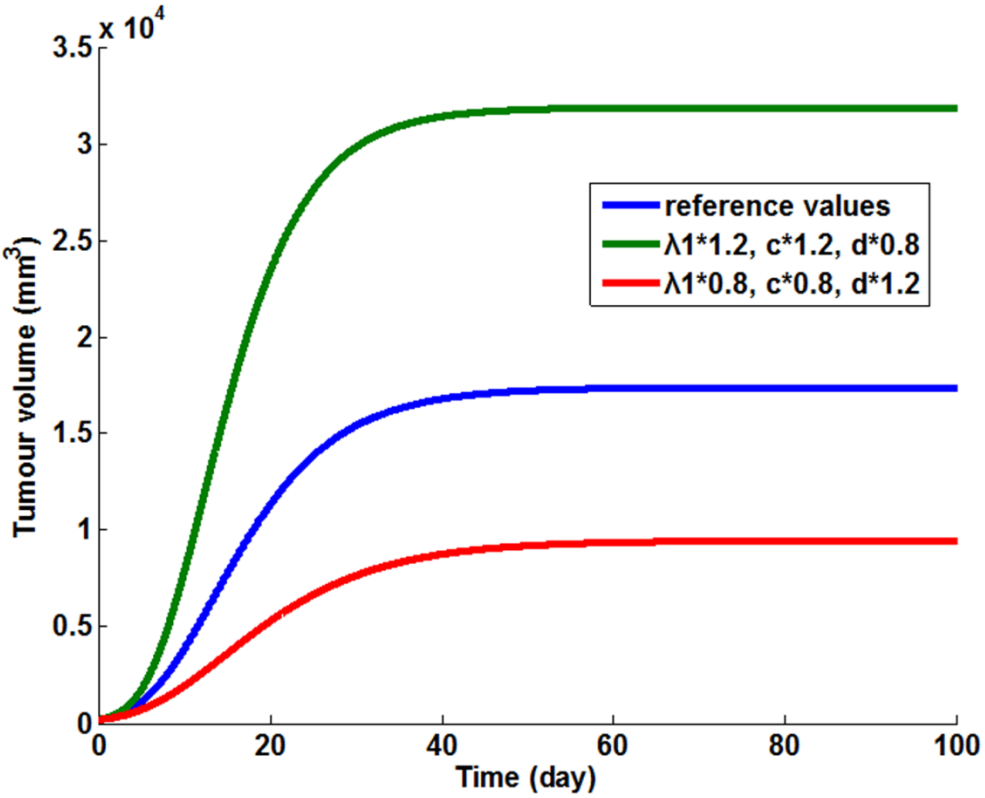


Figure B.5: The combined effect of perturbing λ_1, *c* and *d* by a 20% so as to construct a more aggressive tumour (green line) and a less aggressive tumour (red line) respectively.

Table B.4: Combined effect of varying the input parameters value on the value of the tumour plateau as well as on the time-point of its attainment

Combination of parameters	Variation	Time of plateau (days)	Value of plateau (mm^3)	Variation of plateau value (%)
baseline	$(\lambda_1^{ref}, c^{ref}, d^{ref})$	76	17346	n/a
(λ_1, *c*)	$(1.2\cdot\lambda_1^{ref}, 1.2\cdot c^{ref})$	78	22802	+31.45%
	$(0.8\cdot\lambda_1^{ref}, 0.8\cdot c^{ref})$	112	12412	-28.44%
(λ_1, *d*)	$(1.2\cdot\lambda_1^{ref}, 0.8\cdot d^{ref})$	78	24242	+39.76%
	$(0.8\cdot\lambda_1^{ref}, 1.2\cdot d^{ref})$	116	13196	-23.92%
(*c*, *d*)	$(1.2\cdot c^{ref}, 0.8\cdot d^{ref})$	95	31866	**+83.71%**
	$(0.8\cdot c^{ref}, 1.2\cdot d^{ref})$	94	9442	**-45.57%**
(λ_1, *c*, *d*)	$(1.2\cdot\lambda_1^{ref}, 1.2\cdot c^{ref}, 0.8\cdot d^{ref})$	80	31867	+83.71%
	$(0.8\cdot\lambda_1^{ref}, 0.8\cdot c^{ref}, 1.2\cdot d^{ref})$	102	9441	-45.57%

B.3 Local fitting of the model: reproducing a series of *in vivo* experiments in mice treated with bevacizumab

The next step following the development of the previously outlined continuum approach describing vascular tumour growth under angiogenic signaling and its extension via the inclusion of bevacizumab pharmacokinetics further strengthens the fact that the mathematical structure of the model reflects the nature of the problem under study. To this end, we proceed with the reproduction of a series of *in vivo* experiments in mice treated with bevacizumab. Eight *in vivo* experiments conducted in mice bearing breast, lung (H226), head and neck (SCC1) as well as colon (HCT116, HT29, HCP40 and HP40) tumour xenografts have been selected through relevant literature (Higgins *et al*., 2007; Hoang *et al*., 2012; Selvakumaran *et al*., 2008) in order to fit the specific modelling approach to actual experimental data *in vivo*. These experiments have aimed at examining the anti-tumour activity of bevacizumab mono-therapy and combination treatment.

For the needs of the adaptation of the model to actual experimental data, the code has been properly adjusted via the implementation of two extra m-files. Specifically, a cost function has been defined as the difference of the model prediction from the actual value for the available instances of experimental data. The goal of the adaptation process is to minimize the objective function in order to obtain the most accurate estimates of the variables under study. Hence, a script file has been implemented so as to specify the parameter values that minimize the squared value of the objective function of each data-point using the Matlab function lsqnonlin.

In order to quantify the experimental data obtained through relevant literature (Higgins *et al*., 2007; Hoang *et al*., 2012; Selvakumaran *et al*., 2008), the PlotDigitizer software has been utilize. The latter allows the extraction of data (mean value ± standard deviation) via its digitization.

Starting from the reference input parameter values stated in (Hahnfeldt *et al.*, 1999; Poleszczuk *et al*., 2011), which stem from thorough and extended fitting of the basic core of the model to experimental data from anti-angiogenically treated and untreated Lewis lung tumours in mice, local fitting has been performed with respect to the four non-zero parameters involved in the vascular tumour growth model (λ_1, c, d, β) while in accordance to (Poleszczuk *et al*., 2011), we set α=1. The values of the parameters involved in the two – compartmental pharmacokinetic model are adopted from (Wu *et al*., 2012) and are also only applicable to the case of host – mouse.

While it has been originally expected to determine one quadruple per experiment, we have ended up with one quadruple per experimental curve, a fact that is probably related to the heterogeneity among animal groups consisting of a small number of mice. In general, a good agreement of the simulations with experimental data has been observed, a fact that supports the validity of the underlying model. All results have been obtained with a maximum termination tolerance on the function value equal to 10^{-6}.

Questions have emerged about a few cases where a not satisfying fit has been obtained. Among possible causes could have been human error in digitization of the experimental data by utilizing the PlotDigitizer software and the locality of the selected fitting method. Indeed, local solvers such as lsqnonlin provide the user with a local minimum depending on the selected starting point, but not necessarily the best or global minimum. This means that had the solver been initialized at different parameter values than the selected ones (reference values concerning an untreated tumour (Hahnfeldt *et al.*, 1999; Poleszczuk *et al*., 2011)), better results could have come up. This claim is also supported if one takes into account that the aforementioned fitting difficulty seems to mostly occur

in the therapeutic module of the model and especially in cases of large dosage administration (15 mg/kg, 25 mg/kg) where a greater perturbation of free growth parameter values is expected.

In order to address the aforementioned limited discrepancies, numerous actions need to be taken. First of all, actual clinical data concerning breast cancer patients should be provided by the respective p-medicine partners so as to elucidate as soon as possible the exact reasons of the occurrence of these incidents and act accordingly. For the time being due to the lack of clinical data, extra experimental datasets are extracted from relevant literature in order to corroborate our previous conclusions, risking however the human error due to the digitization of the experimental data with the use of PlotDigitizer software. Last but not least, the global fitting of the model is currently conducted by using the Multistart solver in Global Optimization Toolbox, which runs the selected local solver from multiple starting points, in order to attempt to determine solutions for the problematic cases of fitting encountered so far. Until now, Multistart has been used in conjuction with the Matlab functions lsqnonlin, fmincon and fminsearch.

Description of simulated experiments and local fitting results

Experiment 1: In the context of a mono-therapy antitumour efficacy study, twenty mice bearing KPL-4 human estrogen receptor-negative breast adenocarcinoma xenografts were divided into a vehicle and a treatment group dosed with bevacizumab at 5 mg/kg twice weekly intraperitoneally (Higgins *et al*., 2007*)*. The results of the local fitting are presented in Table B.5 and Fig. B.6.

Experiment 2: In the context of a combination treatment antitumour efficacy study, twenty mice bearing KPL-4 human estrogen receptor-negative breast adenocarcinoma xenografts were divided into a vehicle and a treatment group dosed with bevacizumab at 5 mg/kg twice weekly intraperitoneally (Higgins *et al*., 2007*)*. The results of the local fitting are presented in Table B.5 and Fig.B.6.

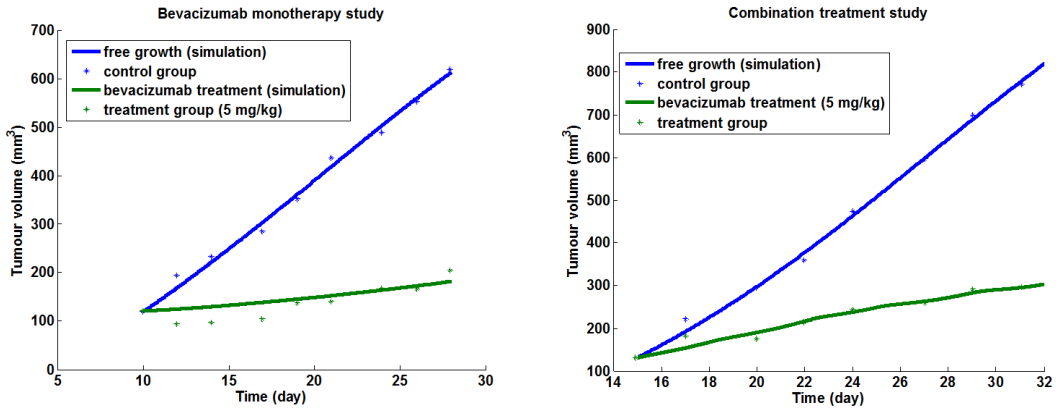


Fig. B.6 Graphical representation of the fitting results of the vascular tumour growth model to the experimental data concerning KPL-4 human estrogen receptor-negative breast adenocarcinoma xenografts that were extracted from (Higgins *et al*., 2008)

TABLE B.5 Parameter values that were specified through locally fitting the vascular tumour growth model to experimental data extracted from (Higgins *et al*., 2007)

STUDY	EXPERIMENTAL GROUP	λ_1	*c*	*d*	α	β
MONO-THERAPY	*CONTROL*	0.1208	0.9329	0.0089	1	1
	TREATMENT (5 mg/kg)	0.0045	29.6758	0.0007	1	32.3360
COMBINATION TREATMENT	*CONTROL*	0.1370	1.0226	0.0078	1	1
	TREATMENT (5 mg/kg)	0.0641	8.5535	0.0290	1	1.3135

Experiment 3: Four groups of athymic mice bearing H226 xenografts (lung cancer cells) were treated with IgG (control) or 3 dose levels of bevacizumab (1, 5 and 25 mg/kg intraperitoneally) twice weekly for a total of 9 doses (3 mice per group) (Hoang *et al*., 2012). The results of the local fitting are presented in Table B.6 and Fig. B.7.

Experiment 4: Four groups of athymic mice bearing SCC1 xenografts (head and neck cancer cells) were treated with IgG (control) or 3 dose levels of bevacizumab (1, 5 and 25 mg/kg intraperitoneally) twice weekly for a total of 9 doses (3 mice per group) (Hoang *et al*., 2012). The results of the local fitting are presented in Table B.6 and Fig. B.8.

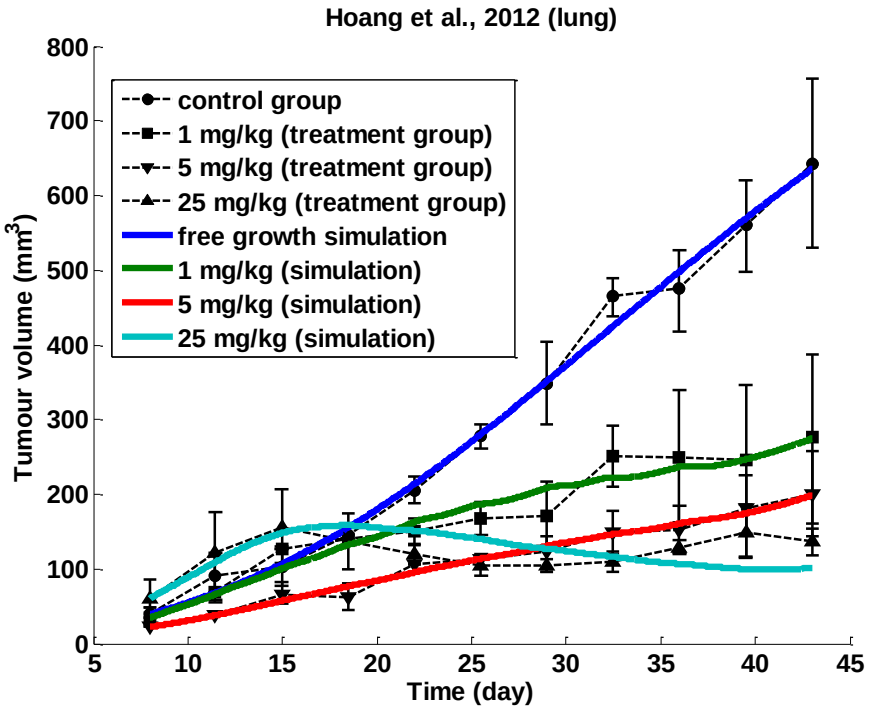


Fig. B.7 Graphical representation of the fitting results of the vascular tumour growth model to the experimental data concerning head and neck tumour xenografts (Selvakumaran *et al.*, 2008)

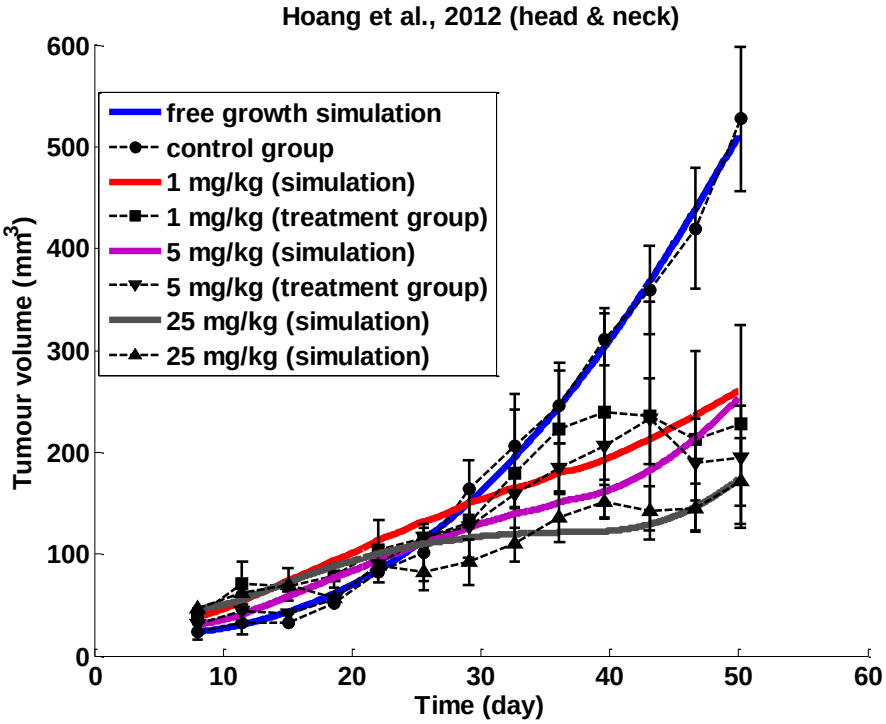


Fig. B.8 Graphical representation of the fitting results of the vascular tumour growth model to the experimental data concerning head and neck tumour xenografts (Selvakumaran *et al.*, 2008)

TABLE B.6 Parameter values that were specified through locally fitting the vascular tumour growth model to experimental data extracted from (Hoang *et al*., 2012)

CELL LINE	EXPERIMENTAL GROUP	λ_1	c	d	α	β
H226 (lung cancer cells)	*CONTROL*	0.0986	1.3031	0.0031	1	1
	TREATMENT 1 (1 mg/kg)	0.048	5.7895	0	1	1.0917
	TREATMENT 2 (5 mg/kg)	0.0651	1.8368	0.0024	1	1
	TREATMENT 3 (25 mg/kg)	0.9281	9.4755	0.0761	1	1.122
HSCC1 (head and neck cancer cells)	*CONTROL*	0.0721	2.2437	0.0036	1	1
	TREATMENT 1 (1 mg/kg)	0.0394	4.9097	0	1	2.5541
	TREATMENT 2 (5 mg/kg)	0.111	1.4957	0.0045	1	1
	TREATMENT 3 (25 mg/kg)	0.0801	4.5639	0.0006	1	1.2544

Experiment 5: Xenograft mouse tumour models of colon cancer cell lines HCT116 were divided in a control and a treatment group. The treatment group received 15 mg/kg of bevacizumab intraperitoneally twice weekly for four weeks, while the control group received the same amount of nonspecific murine IgG antibody (Selvakumaran *et al*., 2008). The results of the local fitting are presented in Table B.7 and Fig. B.9.

Experiment 6: Xenograft mouse tumour models of colon cancer cell lines HT29 were divided in a control and a treatment group. The treatment group received 15 mg/kg of bevacizumab intraperitoneally twice weekly, while the control group received the same amount of nonspecific murine IgG antibody (Selvakumaran *et al*., 2008). The results of the local fitting are presented in Table B.7 and Fig. B.9.

Experiment 7: Xenograft mouse tumour models of colon cancer cell lines HCP40 were divided in a control and a treatment group. The treatment group received 15 mg/kg of bevacizumab intraperitoneally twice weekly, while the control group received the same amount of nonspecific murine IgG antibody (Selvakumaran *et al*., 2008). The results of the local fitting are presented in Table B.7 and Fig. B.9.

Experiment 8: Xenograft mouse tumour models of colon cancer cell lines HP40 were divided in a control and a treatment group. The treatment group received 15 mg/kg of bevacizumab intraperitoneally twice weekly, while the control group received the same amount of nonspecific murine IgG antibody (Selvakumaran *et al*., 2008). The results of the local fitting are presented in Table B.7 and Fig. B.9.

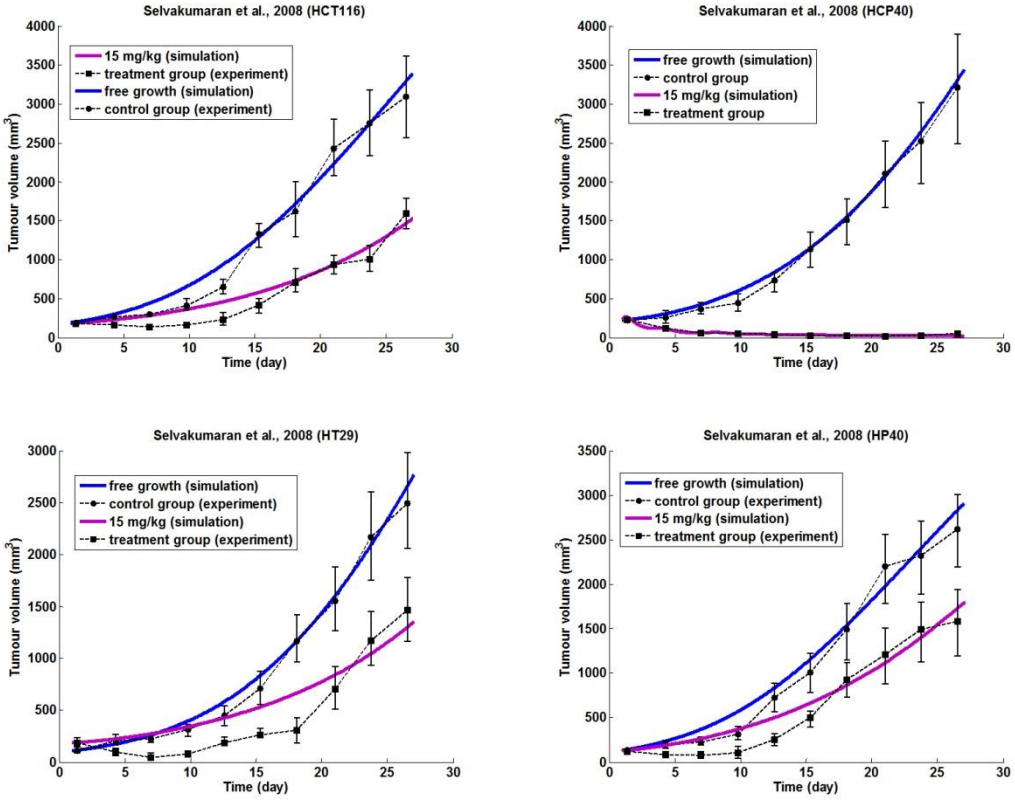


Figure B.9 Graphical representation of the fitting results of the vascular tumour growth model to the experimental data that were extracted to (Selvakumaran *et al.*, 2008)

TABLE B.7 Parameter values that were specified through locally fitting the vascular tumour growth model to experimental data extracted from (Selvakumaran *et al*., 2008)

CELL LINE	EXPERIMENTAL GROUP	λ_1	*c*	*d*	α	β
HCT116	*CONTROL*	0.0986	1.3031	0.0031	1	1
	TREATMENT (15 mg/kg)	0.048	5.7895	0	1	1.0917
HCP40	*CONTROL*	0.0651	1.8368	0.0024	1	1
	TREATMENT (15 mg/kg)	0.9281	9.4755	0.0761	1	1.122
HT29	*CONTROL*	0.0721	2.2437	0.0036	1	1
	TREATMENT (15 mg/kg)	0.0394	4.9097	0	1	2.5541
HP40	*CONTROL*	0.111	1.4957	0.0045	1	1
	TREATMENT (15 mg/kg)	0.0801	4.5639	0.0006	1	1.2544

B.4 REFERENCES

Argyri, K.D.; Dionysiou, D.D.; Stamatakos, G.S., "Modelling the interplay between pathological angiogenesis and solid tumour growth: The anti-angiogenic treatment effect," Advanced Research Workshop on In Silico Oncology and Cancer Investigation - The TUMOUR Project Workshop (IARWISOCI), 2012 5th International , vol., no., pp.1,4, 22-23 Oct. 2012

Bertrand J, Mentré F. Mathematical Expressions of the Pharmacokinetic and Pharmacodynamic Models implemented in the MONOLIX software (http://www.monolix.org). 2008.

D.D.Dionysiou, G.S. Stamatakos, N.K.Uzunoglu, K.S.Nikita and A. Marioli , "A Four Dimensional In Vivo Model of Tumour Response to Radiotherapy: Parametric Validation Considering Radiosensitivity, Genetic Profile and Fractionation", J. theor. Biol., vol. 230, pp. 1-20, 2004.

P. Hahnfeldt, D. Panigrahy, J. Folkman and L. Hlatky, Tumour development under angiogenic signaling: A dynamical theory of tumour growth, treatment response and postvascular dormancy, Cancer Res., **59** (1999), 4770-4775.

Higgins B, Kolinsky K, Linn M, et al. Antitumour activity of capecitabine and bevacizumab combination in a human estrogen receptor-negative breast adenocarcinoma xenograft model. Anticancer Res 2007;27:2279–88

Hoang T, Huang S, Armstrong E, Eickhoff JC, Harari PM: Enhancement of radiation response with bevacizumab. J Exp Clin Cancer Res 2012, 31 (37)

Poleszczuk J, Bodnar M, Foryś U., New approach to modelling of antiangiogenic treatment on the basis of Hahnfeldt et al. model, Math Biosci Eng. 2011 Apr;8(2):591-603. doi: 10.3934/mbe.2011.8.591.

G.S.Stamatakos, E.A.Kolokotroni, D.D.Dionysiou, E.Ch.Georgiadi, C.Desmedt. An advanced discrete state - discrete event multiscale simulation model of the response of a solid tumour to chemotherapy: Mimicking a clinical study. Journal of Theoretical Biology 266, 124-139, 2010

M. Selvakumaran, K.S. Yao, M.D. Feldman *et al.,* Antitumour effect of the angiogenesis inhibitor bevacizumab is dependent on susceptibility of tumours to hypoxia-induced apoptosis, Biochem Pharmacol, 75 (2008), pp. 627–638

Wu F, Tamhane M, Morris M. Pharmacokinetics, Lymph Node Uptake, and Mechanistic PK Model of Near-Infrared Dye-Labeled Bevacizumab After IV and SC Administration in Mice. The AAPS Journal 2012; 14: 252-261

L. Report on the clinical adaptation and optimization of the acute lymphoblastic leukemia (ALL) Oncosimulator models. *In silico* studies

L.1 ALL: A preliminary study exploiting clinical trial data for the clinical adaptation of the core model

L.1.1 Introduction

In the present chapter a preliminary clinical adaptation study of the ALL branch of the *Oncosimulator* is presented. The study is focused on the cytotoxic efficacy of prednisone, as expressed by the apparent *cell kill rate* (CKR) of the drug, i.e. the fraction of cancer cells that are lethally hit by the drug after each chemotherapeutic session. The CKR can be thought of as summarizing important genetic determinants influencing the tumour response to prednisone prephase therapy. Here, a plausible value of the apparent *cell kill rate* (CKR) of prednisone, one of the model input parameters, is suggested following the exploitation of the actual clinical data and the simulation itself. In section L.2 the CKR estimates derived from the present study are correlated with the whole genome expression data of each patient. The aim is to end up with a regression model able to predict the most appropriate value for the CKR parameter of the Oncosimulator, based on the patient's gene expression levels at the time of diagnosis.

A detailed description of the core model (Fig. L.1.1), developed for the simulation of the temporal evolution and response to therapy of acute lymphoblastic leukemia (ALL), can be found in deliverable D12.3.

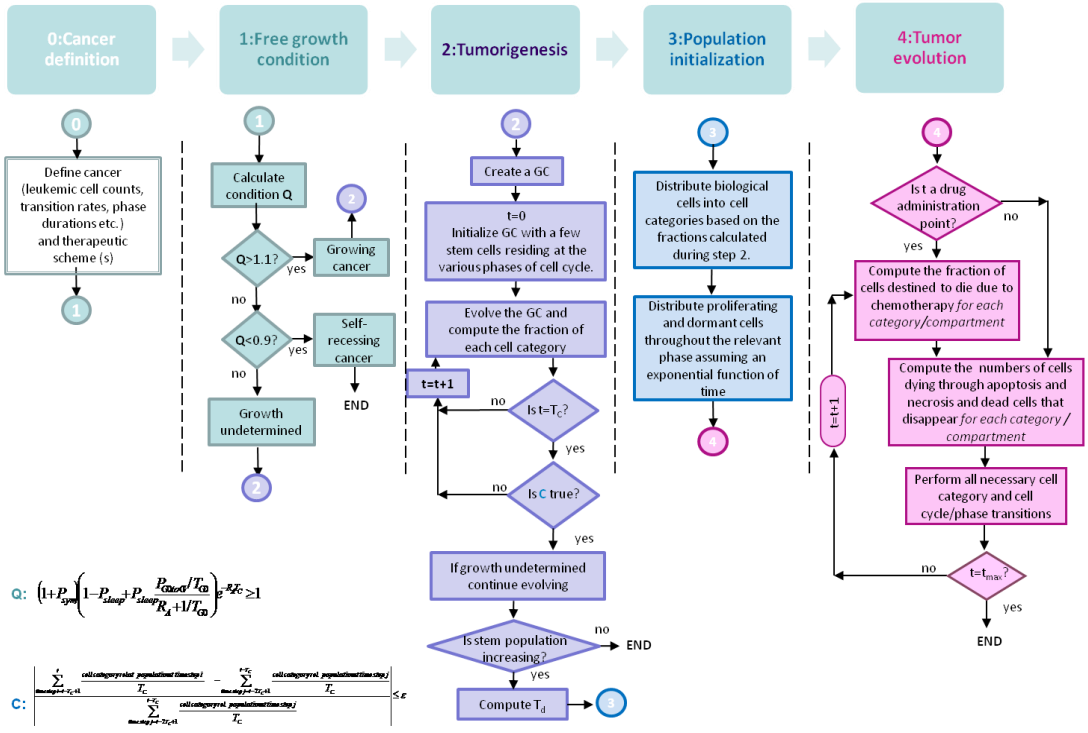


Fig. L.1.1: The *Oncosimulator* workflow

L.1.2 Clinical data

In the present study the ALL Branch of the *Oncosimulator* is applied to a set of multiscale data originating from 191 patients enrolled in ALL BFM 2000 clinical trial. The anonymized data has been provided by the Pediatric Department of the University Hospital Schleswig-Holstein (UK-SH) of Kiel (CAU), Germany. The data provided include:

- **Basic clinical data**: gender, patient/family medical history, age, weight, length, general condition, liver and spleen dimensions
- **Disease related data at diagnosis**: white blood cell count, blood blasts percentage, bone marrow blasts percentage, hemoglobin levels and platelet counts, LDH level, immunological classification, myeloid coexpression, CD10 expression, TdT protein level, DNA index, cytogenetic aberrations, chromosomal translocations (TEL/AML1, BCR/ABL, MLL/AF4), number of sensitive markers, extramedullary disease (CNS, testis, and others).
- **Treatment data**: risk group stratification, CNS therapy
- **Response data**: prednisone response, white blood cell count on day 8, blood blasts percentage on day 8, bone marrow blast percentage on day 15, MRD on days 33 and 78, MRD classification, liver and spleen on day 8, remission at day 33, complete remission
- **Follow up data**: event free survival, relapse, secondary malignancy, stem cell transplantation, survival, mortality
- **Gene expression data**: whole genome expression data

A clinical case is exploitable by the *Oncosimulator*, if it fulfills the following criteria:

- **Leukemic cell counts :**
  - **Minimum requirement**: The *same cell count quantity and the actual date of measurement* should be provided for at least two time instances, one before and, preferably, one after the completion of therapy
  - **Optimal:** The *same quantity and the time of measurement* to be provided for additional time instances, before, during or after the completion of therapy
- **Treatment data:**
  - **Chemotherapy:** *Drugs, Dose* and *Administration time points* (actual dates or in respect to a time reference-the same time reference should be used with the cell count measurements)
  - **Irradiation** (if any): *Dose* and *Administration time points* (actual dates or in respect to a time reference-the same time reference should be used with the cell count measurements)
  - Additional therapy (Hematopoietic stem cell transplantation, other?)
- **Immunophenotype/cytogenetic/molecular data** that could allow us an estimation of leukemic cell kinetics and drug sensitivity/resistance

Fig. L.1.2 summarizes the exploitable clinical data related to treatment and leukemic cell counts for ALL BFM clinical cases.

The proof of concept study presented later in this chapter focuses on the prednisone prephase part of treatment. In order to evaluate the *in vivo* sensitivity of peripheral blood leukemic blasts to prednisone-induced apoptosis for stratification purposes, patients receive a prephase prednisone therapy consisting of a 7 day monotherapy with prednisone (60 mg/m^2) and one age-adjusted dose of intrathecal methotrexate on day 1, that falls in the category of low-dose MTX treatment. The study aims to adapt the Oncosimulator based on the reduction of leukemic blast counts between diagnosis and day 8. We assume that the absolute number of peripheral blood leukemic blasts (/mm^3) is the product of white blood cell (WBC) count (/mm^3) and the peripheral blood (PB) blast percentage. Furthermore, since the time lapsed between diagnosis and treatment onset is not known we assume that the cell counts at diagnosis match those at the beginning of treatment. From

the cohort of the 191 patients, 33 patients miss a peripheral blood cell count either at diagnosis or at day 8 and, hence, 158 patients fulfill the criteria to be included in the adaptation study.

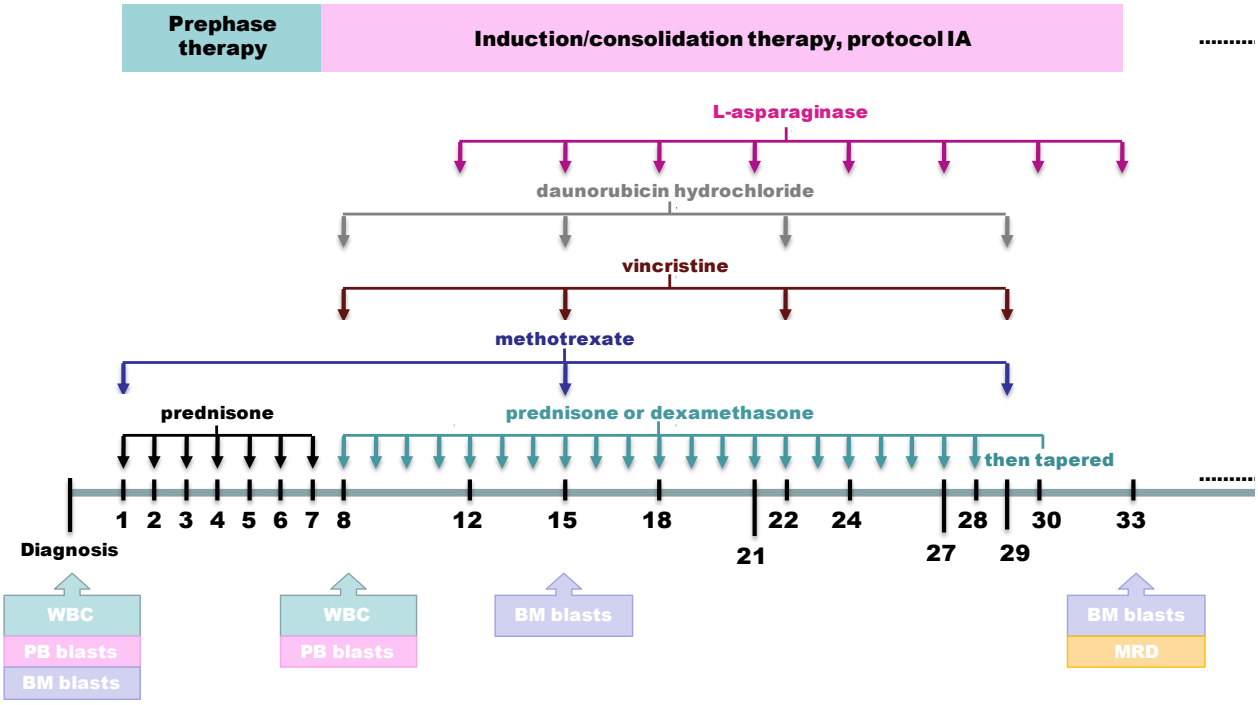


Fig. L.1.2: ALL BFM treatment protocol and time points of leukemic cell counts measurements

L.1.3 The cytokinetic model

In the case of acute lymphoblastic leukemia (ALL) three types of cancer cell compartments can be distinguished: the bone marrow, the circulating blood and the extramedullary blast cell compartments. Due to lack of exploitable data (cell counts) regarding the extramedullary compartment, the cytokinetic model adopted here, considers the compartment of cancer origin, which is the bone marrow cell compartment and the circulating blood.

L.1.3.1 ALL and the cancer stem cell theory

Three major models have been described in literature for tumour propagation: the stochastic model, the clonal evolution model and the cancer stem cell model. The stochastic model predicts that all tumour cells are biologically equivalent, their behaviour is subject to stochastic influences by intrinsic or extrinsic factors and they can revert from one state to another because these influences do not induce permanent changes (Dick, 2008). The clonal evolution model proposes that neoplasms arise from a single cell of origin, and tumour progression results from acquired genetic variability within the original clone allowing sequential selection of more aggressive sublines (Nowell, 1976). The cancer stem cell model is based on the hypothesis that tumours are organized in a hierarchical way, and only a small proportion of cells with stem-like properties has the capacity for self-renewal, multilineage differentiation, tumour sustenance and formation of new tumours (Reya et al., 2001). The origin of the cancer stem cell is not necessarily the normal stem cell but could be a committed progenitor cell that reverts to a stem-like phenotype during transformation.

Even though cancer stem cells have been identified in a wide range of human malignancies, including acute myeloid leukemia, published data on ALL stem cells are conflicting (Bernt and Armstrong, 2009; Reaman and Smith, 2011). The debate concerns both the cell of origin, namely whether it is a lymphoid progenitor or a more primitive pluripotent hematopoietic stem cell, and the hierarchical organization of ALL, with some studies suggesting that lymphoid progenitors may not lose their self-renewal ability with differentiation. Fig. L.1.3 summarizes some of the models proposed in literature for ALL.

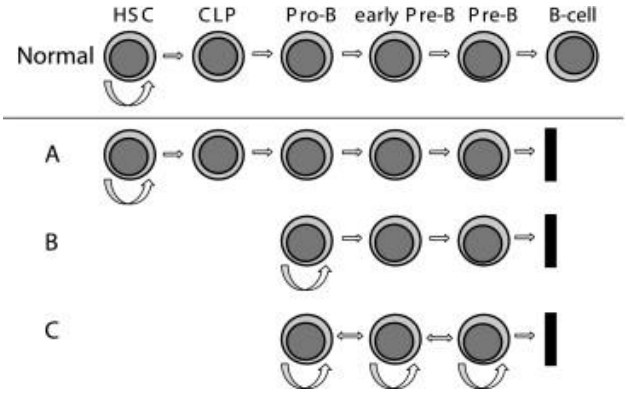


Fig. L.1.3: Hierarchy and differentiation in different leukemia stem cell models. Normal Differentiation toward the B-cell lineage is shown. (A) Leukemia sustained by a very early leukemia stem cell (similar to a hematopoietic stem cell or very early progenitor). The leukemic cells continue to develop along the lymphoid lineage and arrest at the pre-B stage. (B) Leukemia sustained by a leukemia stem cell at the developmental stage of a committed progenitor (here a pro-B cell), with further differentiation of its progeny along the lymphoid lineage, and arrest at the pre-B stage. (C) Leukemia maintained by "leukemia stem cells" corresponding to different stages in early B-cell development. Multiple subpopulations possess the ability to self-renew and propagate the leukemia. (Figure taken from Bernt and Armstrong, 2009.)

Due to the lack of concluding evidence for the exact hierarchy (or absence) within the leukemic clone and the frequency of leukemic stem cells in the different ALL subpopulations, the cytokinetic model adopted for the purposes of ALL propagation modelling is based on the well documented generic hypothesis of cancer stem cell theory. However the hypothesis that more mature cell populations may possess the ability to self-renew will also be examined within the last year of P-medicine.

L.1.3.2 Cell kinetics of ALL

Lymphoid malignancies of precursor B cell and T cell origin seem to have a longer cell cycle than normal bone marrow precursor cells, even though the duration of DNA synthesis phase seems to be identical. Indicative values of the cell cycle duration of leukemic cells vary considerably, taking value between 25-240 hours (Cooperman et al., 2004; Hirt et al., 1992; Tsurusawa et al., 1992). The length of G1 phase is believed to be highly variable and can range from about 6 hours to several days or longer (Kufe et al., 2003). The S-phase duration (Ts) seems to show much less variation within and between tumours, and between tumours and normal tissues, varying between 10–20 hours. (Rew and Wilson, 2000). The duration of mitosis phase is considered constant and equal to 1h (Bast et al., 2000). The duration of cell cycle phases is used to determine the time point that hit cells die through apoptosis depending on the action mechanism of the drug.

In most cases, the proportion of leukemic cells in S phase is lower than normal cells (LI), even though the growth fraction (GF) can be high. In the study of Hirt, 1992, this controversy has been explained by a prolong stay of cells in G1 phase. The authors conclude that leukemic cells from children with precursor B or T-cell ALL are continuously cycling without leaving the cell cycle for a rest phase.

Apoptosis, from initiation to cell elimination, is a relatively short process, with a mean duration of about 1-6h (Hirt et al., 1997; Gavrieli *et al*., 1992; Bursch et al. 1990). In consistency, apoptotic cells seems to be also rapidly removed from the circulation as a result of phatocytosis (< 3 hours unless phagocytes capacity is exceeded) (Groninger et al., 2000). In contrast, the removal of necrosis products is a time consuming process requiring even several days to be completed (Ginsberg T., 1996).

The doubling time of acute leukemia can be very rapid, even 4 days, in relapsing children. In the study of Skipper & Perry (1970), the authors argue that the growth rate can slow down to about 27 days in advanced stages of the disease, an estimate based on the mean survival time of untreated children. In the study of Hirt et al. (2011), the authors have examined the proliferation of $CD19^+$ B-cell precursor ALL and have identified two distinct leukemic blast populations with mean doubling times equal to 10.7 and 294 days respectively.

TABLE L.1.1 Cell cycle phases duration for leukemic and normal hematopoietic cells: Bibliographic review

Study		T_S (h) (Range)	T_C (h) (Range)
	Leukemic cells		
Hirt *et al.*, 1992	Early pre-B ALL(n=19)	18.6±2.8[a]	116.4±55.9[a]
Hirt *et al.*, 1992	Pre-B ALL (n=13)	19.1±1.9[a]	112.5±46.8[a]
Hirt *et al.*, 1992	T ALL/NHL (n=8)	19.9±2.5[a]	102.1±22.6[a]
Hirt *et al.*, 1992	B ALL/NHL (n=6)	9.6±0.9[a]	24.9±8.1[a]
Tsurusava *et al.*, 1992		14.3[b] (6.1-34.4)	T/B ALL: 34 Early B ALL: 75.2
Cooperman et al., 2004	B cells		(26-240)
Beesley *et al.*, 2006	T ALL cell lines (n=15)		(23-442)
Beesley *et al.*, 2006	Pre-B ALL cell lines (n=5)		(53-108)
Beesley *et al.*, 2006	B ALL cell lines(n=2)		53 and 96
	Normal cells		
Hirt *et al.*, 1992	pre-B (n=7)	18.3±1.8[a]	65.5±3.5[a]
Cooperman et al., 2004	precursor B $CD19^+C10^-$		90.5
Cooperman et al., 2004	precursor B $CD19^+C10^-$		No cell division

[a]Mean±SD
[b]Median

TABLE L.1.2 Proliferation indexes for leukemic and normal hematopoietic cells: Bibliographic review

Study		GF (%) (Range)	LI (%) (Range)	Apoptic cells (%)	Resting G0 (Range)	Td (days) (Range)
			Leukemic cells			
Hirt *et al.*, 1992	Early pre-B ALL(n=19)	38.7±11.2	7.3±3.7 [a]			
Hirt *et al.*, 1992	Pre-B ALL (n=13)	46.9±16.1	9.1±5.2			
Hirt *et al.*, 1992	T ALL/NHL (n=8)	47.9±14	9.9±4 [a]			
Hirt *et al.*, 1992	B ALL/NHL (n=6)	86.5±13.7	36.8±13.3 [a]			
Hirt *et al.*, 1997a	Pre-B and T ALL	PCNA : 83.6±16.1 [a] (30-96) P120: 94.1±2.6 [a] (89-98)			2.9±3.8 [a] (0.5-18)	
Hirt *et al.*, 1997b	Precursor B ALL(n=23)		8.2±6.3 [a] (0.9-22.4)	0.9±1.4 [a] (0.1-7.2)		
Gorczyca et al. 1993	Blood ALL, AML, CML (n=11)			(1-8)		
Groninger *et al.*, 2000	Blood ALL (n=5)			(0-2)		
Cheung *et al.*, 1972	ALL(n=11)		BM:11.2±6.2 [a] (2-22.6) PB: 8.8±8.4 [a] (1.4-28)			
Inoue *et al.*, 1993	AL (n=32)	PB:12% BM:15%	PB:5.6% BM:10.1%			
Schultz *et al.*, 1989	ALL (n=1)					6.5
Skipper & Perry, 1970	From studies of ALL at beginning of relapse		PB:0.7-6.2 BM:2.6-15			StudyI: 4 StudyII: 5.8 (2-16)
Hirt et al. 2011	$CD19^+$ B-cell precursor ALL (n=15)		$CD19^+$: 7.2 [b] (5.7-8.8) $CD19^-$: 0.19 [b] (0.15-0.4)			$CD19^+$: 10.7 [b] (8.2-14.9) $CD19^-$: 294 [b] (135-354)
			Normal cells			
Cheung *et al.*, 1972			26.9±6.7 [a] (18-40)			
Hirt *et al.*, 1992	pre-B (n=7)	66.9±7.3	18.7±1.9 [a]			

[a]Mean±SD/ [b]Median/ PB: Peripheral blood/ BM: Bone marrow

L.1.3.4 Modelling of cancer growth and response to chemotherapy

The adopted cytokinetic model (Fig. L.1.4) incorporates the progression through the active cell cycle, the 'exit' to the quiescent state, the differentiation pathway, the passage to the blood circulation and cell loss. Cancer progression is sustained by a rare cell population that exhibits stem cell like properties located in the bone marrow compartment. These so called 'cancer stem cells' have the ability to self-renew, as well as to give rise to cells (LIMP cells) that follow the developmental hierarchy towards differentiation. Differentiation is assumed to be arrested at the last considered LIMP stage. Only cells of the last LIMP stage are allowed to leave the bone marrow compartment and enter circulation (peripheral blasts). The passage is implemented through the mitosis phase.

A proliferating cancer cell of the bone marrow compartment, either stem or LIMP, passes through the successive cell cycle phases (G1, S, G2, M). Upon completion of mitotic division, a fraction of daughter cells will become dormant. On top of the internal molecular pathway interactions, the local oxygen and nutrient supply conditions regulate the transition to dormant, G0, phase and the "awakening" of dormant cells. Dormant cells are assumed to survive only for a limited time length. After the expiration of this time, dormant cells die through necrosis, unless the local conditions allow the re-entering of the dormant cells into the active cell cycle. At a first approximation circulating peripheral blasts are considered quiescent. No recirculation of leukemic blasts from blood to bone marrow is assumed. All living cell categories may die through spontaneous apoptosis.

A fine partitioning of proliferating cells (stem and LIMP) into subclasses is introduced by considering a number of compartments that equals the duration of the cell cycle. Each compartment corresponds to each hour of the total duration of the G1 (gap 1), S (synthesis), G2 (gap 2), M (mitosis) phases and therefore contains the biological cells residing in the given hour. Similarly dormant cells (stem and LIMP) are distributed in a number of compartments that correspond to each hour of the duration of the G0 phase.

When cancer is chemotherapeutically treated, a fraction of cancer cells are lethally hit by the drug or its metabolites. Lethally hit cycling cancer cells of the bone marrow enter a rudimentary cell cycle that leads to apoptotic death via a specific phase dictated by the action mechanism of the chemotherapeutic agent used. Similarly, in the case of cell cycle non specific drugs, lethally hit dormant (G0) cells enter the G0hit phase. Marking of a cell as hit by the drug is assumed to take place at the instant of drug administration although its actual time of death is dictated by the specific pharmacokinetics and pharmacodynamics of the drug. It is pointed out however that cell cycle phase specific drugs can be readily modeled by the cytokinetic model shown in Figure L.1.4
Fig. by appropriately selecting the "Chemo" induced exit from the normal cell cycle for both cases of stem and LIMP cells. In blood circulation, an instant death of the lethally hit cancer cells at the time of drug administration is assumed, through the apoptotic machinery.

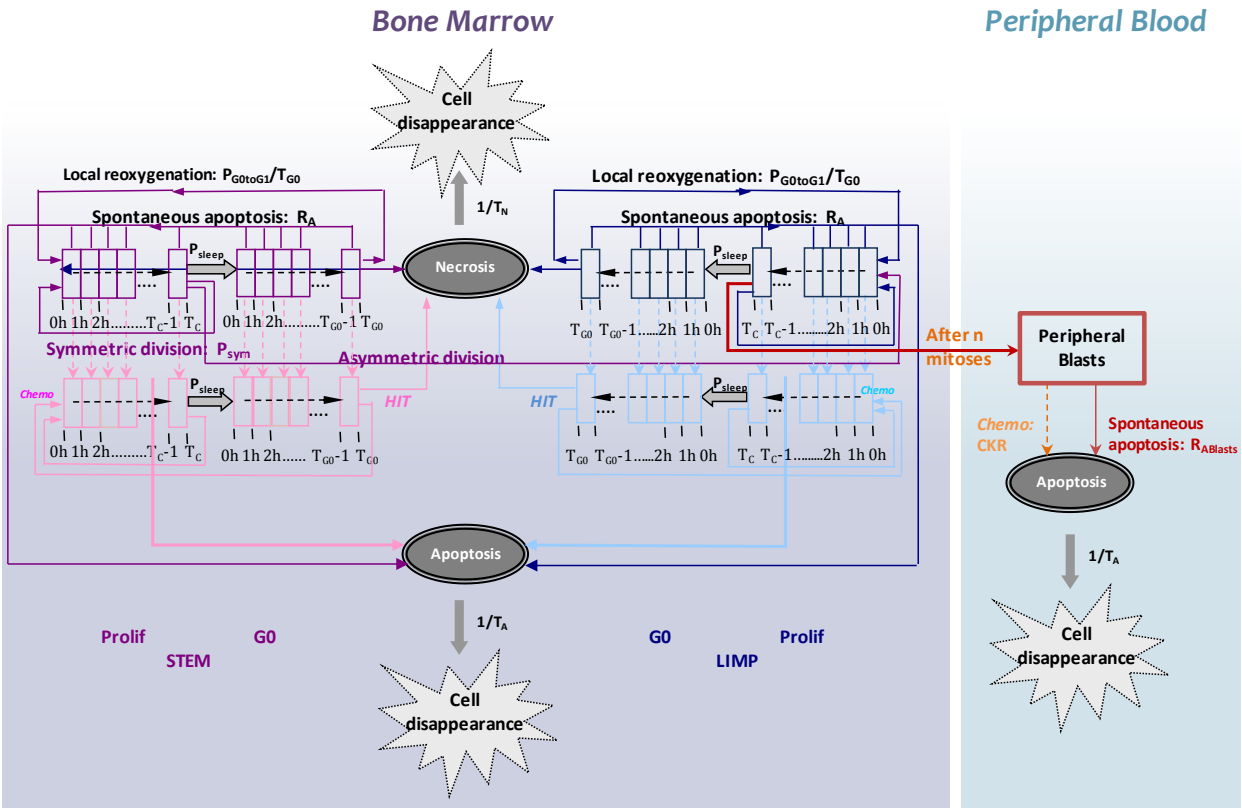


Fig. L1.4 Generic cytokinetic model of bone marrow and peripheral blood cell compartments (cell category / phase transition cell. LIMP: LImited Mitotic Potential cancer cell (also called committed or restricted progenitor cell). Prolif: Proliferating cell. G0: dormant cell. Chemo: chemotherapeutic treatment. HIT: lethally hit cells by the drugdiagram) for cancer response to chemotherapy. Abbreviations: STEM: stem

L.1.3.5 Modelling of the action mechanisms of the drugs considered

L.1.3.5.1 Methotrexate induced cell loss

Methotrexate (MTX) is a folate analogue, showing activity in the S phase of the cell cycle. The effect of MTX on cell cycle and cell viability is dose dependent (Tsurusawa *et al*. 1988). The mechanism of cell death induced by MTX at high and low doses is apoptosis (Genestier *et al*. 1998). In general, MTX acts by inhibiting several enzymes in the folate pathway, necessary for the synthesis of purines and thymidines (the precursors of nucleotides, the building blocks of DNA). The major mechanism of action involves the inhibition of dihydrofolate reductase (DHFR), resulting in the compromise of dihydrofolate to tetrahydrofolate reduction. MTX depletion of reduced folates blocks synthesis of purine and thymidine, thereby stopping replication and leading to cell death (Taylor and Tattersall 1981; Schmiegelow 2009). Polyglutamation of MTX, increases the anti-proliferative effects of MTX (Walling, 2006; Schmiegelow, 2009) by making it difficult for the cell to transport the MTX out of the cell and by inhibiting folate-dependent enzymes involved in DNA replication (DHFR, thymidylate synthetase (TS), and aminoimidazolecarboxamide ribonucleotide (AICAR) formyltrransferase).

Even though the cytotoxicity of high doses of MTX is primarily due to inhibition of purine and pyrimidine synthesis (Schmiegelow, 2009), the mechanisms of immunosuppressive properties associated with low-dose MTX treatments are not fully elucidated (Genestier *et al*., 1998; Schmiegelow, 2009). However, treatment of peripheral blood with low-dose MTX induced apoptosis of *in vitro* activated T cells (Genestier *et al*., 1998), mainly resulting from DHFR and TS inhibition.

Moreover, in this study the authors argued that the apoptotic signal was most likely initiated during the S phase and showed that the resting T cells were resistant to MTX induced apoptosis unless subsequently activated, finding that suggests that resting T cells may convert MTX to polyglutamate forms and retain them intracellularly for several days. In the study of Tsurusawa *et al*. (1988) doses of 10^{-7} and 10^{-6}M of MTX (corresponding to low dose MTX therapy) strongly reduced the clonogenicity and prevented the progression of cells at the G1/S boundary and across the S phase, but not in the late S or the other phases. Even though it is believed that S-phase cells are the main target of MTX, some investigators have reported that non-S-phase cells are also susceptible to cytotoxicity of MTX, e.g. G1 cells (Goncharova and Frankfurt, 1976) or cycling cells arrested at the G1/S boundary during treatment with MTX (Tsurusawa *et al*., 1988)

In our modelling approach, cancer cells are assumed to absorb the drug at cycling phases only, whereas apoptotic death of hit cells takes place in the S phase.

L.1.3.5.2 Prednisone induced cell loss

Glucocorticoids (GC) act through binding to GC receptors (GR). In the absence of hormone, GR resides in the cytosol combined with a variety of proteins. A direct mechanism of action involves homo-dimerization of the receptor, translocation via active transport into the nucleus, and binding to specific DNA responsive elements activating gene transcription (Ociepa *et al*., 2010). The mechanism of GC acting upon leukemic cells has not been fully elucidated to date, yet it is assumed that the said mechanism arrests the cells in G1 phase to induce apoptosis and modulate the expression of apoptosis regulating proteins, e.g. through exerting a direct inhibitory effect on Bcl-2 protein or via direct activation of the caspase cascade (including caspase-3) (Ociepa *et al*., 2010). Glucocorticoids also induces apoptosis in nonproliferating cells, triggering signaling pathways that do not evolve changes in cell cycle proteins (King and Cidlowski, 1998).

In our modelling approach, cancer cells are assumed to absorb the drug at dormant and cycling phases, whereas apoptotic death of hit cells takes place in the end of G1 phase.

L.1.4 A preliminary clinical adaptation study

The efficacy of the prephase prednisone scheme strongly depends on both the kinetics of cancer cell proliferation and the pharmacodynamics (as reflected by the value of cell kill rate) of the chemotherapeutic agents used. Ideally direct measurements of the above quantities for each patient, derived from bone marrow aspirates and blood samples, would greatly potentiate the predictive ability of the Oncosimulator. However, the provision of such measurements has not been forseen in the context of BFM-ALL trial. The challenge posed here is how to translate the available patient specific data to model parameter values and how to identify these data at cellular and subcellular level (molecular data, genes expression levels etc) that will allow the individualized estimation of parameter values or at least the most plausible value windows that the parameters lie in.

In the present section a proof of concept adaptation study is presented focusing on the estimation of prednisone cell kill rate for a cohort of 158 patients. The value of the cell kill rate could be thought of as summarizing those important genetic determinants and molecular mechanisms that render leukemia cells resistant or sensitive to therapy. At a first approximation one type of cancer growth will be assumed, identical for all patients. The selection of the untreated tumour growth parameter values will be guided by experimental cell kinetic data found in literature (section L.1.3.2). A multi-variate optimization methodology for the simultaneous adaptation of both cell kinetics and pharmakodynamics model parameters for each patient will be examined in the last year of the

project. The derived prednisone sensitivity/resistance profile of the patient cohort will then be exploited for the quest of correlations between prednisone cell kill rate and gene expression levels (see subsequent sections).

The patient specific data that has been exploited by the model at this stage of the adaptation include the applied chemotherapeutic scheme (drugs and administration time points) and the peripheral blood counts (the white blood cell counts (/mm^3) and the percent of blasts in peripheral blood). The peripheral blood counts were provided for two time instants at diagnosis and on day 8 of treatment (Figure L.1.2).

L.1.4.1 Adaptation methodology

A two - step adaptation process is followed. The first step refers to the adaptation of the model parameters that regulate cancer free growth kinetics (Table L.1.3). In general a large number of virtual tumour implementations (virtual instances) corresponding to certain given or known parameter values exist. Narrowing this large window of possible solutions, ideally to one solution for a specific clinical case, is a critical first step in the adaptation procedure. Due to the non-availability of proliferation indices, such as Ki-67, thymidine labeling or mitotic indexes, and data that could allow the precise estimation of the tumour growth, such as at least two leukemic cell counts before the therapy, a literature survey has provided biologically reasonable values for critical tumour kinetics features (section L.1.3.2). In the adaptation paradigm presented here the following assumptions/constraints were imposed, based on literature (section L.1.3.2) (or logical assumptions supported by basic science or clinical experience in case of lack of literature data):

i. A middle value of cell cycle duration (T_c=90h) (Hirt *et al.*, 1992; Cooperman et al., 2004, Beesley *et al.*, 2006)
ii. A duration of apoptosis (T_A) equal to 6h (Hirt et al., 1997)
iii. A duration of necrosis (T_N) equal to 120h (Ginsberg T., 1996)
iv. A spontaneous apoptosis rate (R_A, $R_{ABlasts}$ of all living leukemic cells equal to0.001^{-1}
v. The fraction of dormant cells that re-enter the cell cycle (P_{G0toG1}) is assumed equal to 0.01
vi. A fraction of stem cells in bone marrow that perform symmetric divisions (P_{sym}) equal to 0.4
vii. The number of mitoses performed by LIMP cells before they are arrested (N_{LIMP}) is assumed equal to 7.
viii. A relatively slowly growing cancer with volume doubling time (T_d) ~30 days, since all cases are newly diagnosed and not in recurrence
ix. A high value of growth fraction in bone marrow (~80%)
x. A fraction of dormant cells in bone marrow complementary to growth fraction

The desired volume doubling time was achieved by the adaptation of the parameter that regulates the quiescence of newborn cells after mitosis (P_{sleep}), whereas the relative distribution of leukemic cells in the dormant and the proliferating phases (constrains ix, x) was achieved by the adaptation of the duration of dormant phase (T_{G0}) (Table L.1.3).

At the second step, the cell kill rate (CKR) of the drugs is adapted to the observed blast count reduction. Since the regimen given consists of two chemotherapeutic agents, it is not possible to accurately determine the cell kill rate of each drug from the data provided, even in the ideal case of the availability of all required proliferation indices and tumour free growth kinetics features that would enable an excellent fitting of the model parameters to the clinical case examined. In the paradigm presented here an arbitrarily fixed value for methotrexate CKR (=0.2) has been considered for all patients. The cytotoxic effect of MTX is assumed relatively low since the dosage falls in the category of low-dose therapy.

The adaptation of CKR was performed automatically using an optimization procedure (the fzero command in Matlab), that returns the zero of the distance between the observed blast count and the simulation outcome. The algorithm uses a combination of bisection, secant, and inverse quadratic interpolation methods.

TABLE L.1.3. ADAPTATION OF PARAMETER VALUES USED FOR THE CLINICAL CASES SIMULATIONS AND RESULTING TUMOUR CHARACTERISTICS

Parameter	Description	Value
T_c	Cell cycle duration	90 h
T_{G0}	Duration of dormant phase	96 h
T_N	Time needed for necrosis products to disappear from bone marrow	120 h
T_A	Time needed for apoptosis products to be removed from bone marrow	6 h
N_{LIMP}	Number of mitoses performed by LIMP cells before they are arrested	7
R_A	Apoptosis rate of stem and LIMP cells	0.001 h^{-1}
R_{ABlast}	Apoptosis rate of peripheral blast cells	0.001 h^{-1}
P_{G0toG1}	Fraction of dormant cells that re-enter the cell cycle	0.01
P_{sleep}	Fraction of cells that enter the G0 phase following mitosis	0.159
P_{sym}	Fraction of stem cells that perform symmetric division	0.4
CKR_{MTX}	Cell kill rate of methotrexate	0.2
Tumour Characteristic		**Value**
Doubling Time		35 d
Fraction of Dormant Cells in Bone Marrow		0.17
Fraction of Necrosis in Bone Marrow		0.04
Growth Fraction in Bone Marrow		0.83

L.1.4.2 Results

In figure L.1.5 the frequency of the estimated values of prednisone CKR among the clinical cases are presented. The values of CKR have been grouped in intervals of length = 0.1. Since the total CKR of prednisone and methotrexate cannot exceed 1, the upper limit of prednisone CKR is 0.8. The results are presented based on patient classification into good (<1000 blasts/μL on day 8) and poor (≥1000 blasts/μL on day 8) prednisone responders. Early response is believed to be a strong predictor of treatment outcome and weights in the choice of treatment protocol.

Based on the results, the values of CKR span in the whole value range. 7 clinical cases (1 good and 6 poor) were characterized by a higher blast count on day 8 of treatment in respect to diagnosis. For those patients the simulation failed to achieve the final blast count, even in the absence of therapy, implying a more rapid growth rate than the assumed. For those patients no valid CKR value could be estimated (CKR=0 in Figure L.1.5). On the other hand, 15 patients achieved complete remission of peripheral blasts, resulting in a very high estimated CKR, close to the upper limit (~0.8). The results successfully demonstrate that poor responders tend to have a low value of CKR in contrast to good responders. However, even though poor responders have a clearly lower median value of prednisone CKR than good responders (0.32 compared to 0.59 respectively) the range within CKR fluctuates is considerable wide for both groups and overlap with each other. As a result patients that seem to have similar response to treatment may not have been classified in the same group. However, due to the small size of the sample examined and the consideration of only one type of cancer growth, no definite conclusion can be drawn as yet.

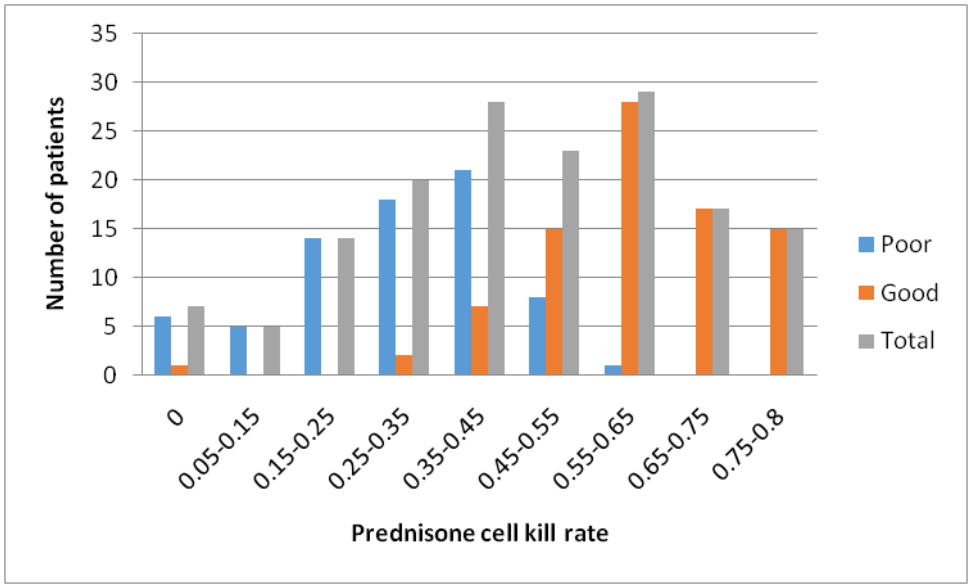


Fig. L1.5: Estimated values of prednisone cell kill rate

L.1.5 REFERENCES

Bast RC, Kufe DW, Pollock RE, Weichelbaum RR, Holland JF, Frei E, *Cancer Medicine* 5th ed. (2000) Decker B.C. Inc, Hamilton, Ontario, Canada.

Beesley AH, Palmer ML, Ford J, Weller RE, Cummings AJ, Freitas JR, Firth MJ, Perera KU, de Klerk NH, Kees UR, Authenticity and drug resistance in a panel of acute lymphoblastic leukaemia cell lines. Br J Cancer. 2006 Dec 4;95(11):1537-44.

Bernt KM and Armstrong SA, 'Leukemia Stem Cells and Human Acute Lymphoblastic Leukemia', Semin Hematol. 2009 Jan;46(1):33-8.

Bursch W, Paffe S, Putz B, Barthel G, Schulte-Hermann R, 'Determination of the length of the histological stages of apoptosis in normal liver and in altered hepatic foci of rats', Carcinogenesis. 1990 May;11(5):847-53.

Cooperman J, Neely R, Teachey DT, Grupp S, Choi JK, 'Cell division rates of primary human precursor B cells in culture reflect in vivo rates', Stem Cells. 2004;22(6):1111-20.

Dick, J. E. (2008). Stem cell concepts renew cancer research. Blood *112*, 4793-4807.

Gavrieli, Y., Sherman, Y. & Ben-Sasson, S.A. (1992) Identification of programmed cell death in situ via specific labeling of nuclear DNA fragmentation. Journal of Cell Biology, 119, 493±500.

Genestier, L., Paillot, R., Fournel, S., Ferraro, C., Miossec, P., Revillard, J.P., 1998. Immunosuppressive properties of methotrexate: apoptosis and clonal deletion of activated peripheral T cells. J. Clin. Invest. 102, 322–328.

Ginsberg T. 'Modellierung und Simulation der Proliferationsregulation und Strahlentherapie normaler und maligner Gewebe'. Fortschritt-Berichte, VDI Verlag: Duesseldorf, Reihe 17: *Biotechnik* (1996), 140:103–107.

Goncharova, S. A., and Frankfurt, O. S. Effect of methotrexate on the cell cycle of L1210 leukemia. Cell Tissue Kinet., 9: 333-340, 1976.

Gorczyca W, Bigman K, Mittelman A, Ahmed T, Gong J, Melamed MR, Darzynkiewicz Z. Induction of DNA strand breaks associated with apoptosis during treatment of leukemias, Leukemia. 1993 May;7(5):659-70.

Groninger E, de Graaf SS, Meeuwsen-de Boer GJ, Sluiter WJ, Poppema S, Vincristine-induced apoptosis in vivo in peripheral blood mononuclear cells of children with acute lymphoblastic leukaemia (ALL), Br J Haematol. 2000 Dec;111(3):875-8.

Hirt A,Werren EM, Luethy AR et al. Cell cycle analysis in lymphoid neoplasia of childhood: differences among immunologic subtypes and similarities in the proliferation of normal and leukaemic precursor B cells. Br J Haematol 1992;80:189–193.

Hirt, A., Antic, V., Wang, E., Lóthy, A.R., Leibundgut, K., Von der Weid, N., Tobler, A., Wagner, H.P., 1997a. Acute lymphoblastic leukaemia in childhood: cell proliferation without rest. Br. J. Haematol. 96, 366–368.

Hirt, A., Leibundgut, K., Lóthy, A.R., Von der Weid, N., Wagner, H.P., 1997b. Cell birth and death in childhood acute lymphoblastic leukaemia: how fast does the neoplastic cell clone expand? Br. J. Haematol. 98, 999–1001.

Howard E. Skipper and Seymour Perry, 'Kinetics of Normal and Leukemic Leukocyte Populations and Relevance to Chemotherapy', *Cancer Res* 1970;30:1883-1897.

Inoue T, Yamane T, Park K, Im T, Tatsumi N, Okuda K, 'Evaluation of proliferating cells in acute leukemia with antibodies to proliferating cell nuclear antigen and Ki-67 monoclonal antibodies.' Arch Pathol Lab Med. 1993 Oct;117(10):969-72.

K. L. King and J. A. Cidlowski, cell cycle regulation and apoptosis Annu. Rev. Physiol. 1998. 60:601–17

Kufe DW, Pollock RE, Weichselbaum RR, et al., editors., Holland-Frei Cancer Medicine. 6th edition. Hamilton (ON): BC Decker; 2003.

Kjeld Schmiegelow, 'Advances in individual prediction of methotrexate toxicity: a review' British Journal of Haematology 2009, 146, 489–503

Nowell PC, "The clonal evolution of tumour cell populations," Science, vol. 194, no. 4260, pp. 23–28, 1976.

Ociepa T, Maloney E, Kamiepska E, Wysocki M, Kurylak A, Matysiak M, Urasipski T, Urasipska E, Domaga³a W., Simultaneous assessment of p53 and MDM2 expression in leukemic cells in response to initial prednisone therapy in children with acute lymphoblastic leukemia., Pol J Pathol. 2010;61(4):199-205.

Gregory H. Reaman and Franklin O. Smith, editors, Childhood Leukemia: A Practical Handbook (Pediatric Oncology), Springer-Verlag Berlin Heidelberg, 2011.

Rew DA and Wilson GD, 'Cell production rates in human tissues and tumours and their significance. Part II: clinical data', European Journal of Surgical Oncology (2000) 26:405–417.

Reya T, Morrison SJ, Clarke MF, and Weissman IL, "Stem cells, cancer, and cancer stem cells," Nature, vol. 414, no. 6859, pp. 105–111, 2001.

Schultz FW, van Dongen JJM, Hahlen K, Hagenbeek A. Time-history of the malignant population in the peripheral blood of a child with T-cell acute lymphoblastic leukemia. Comput Math Applic 18:929, 1989

Taylor IW, Tattersall MH. 'Methotrexate cytotoxicity in cultured human leukemic cells studied by flow cytometry.' Cancer Res. 1981 Apr; 41(4):1549-58.

Tsurusawa M, Niwa M, Katano N, Fujimoto T. 'Flow cytometric analysis by bromodeoxyuridine/DNA assay of cell cycle perturbation of methotrexate-treated mouse L1210 leukemia cells.' Cancer Res. 1988 Aug 1;48(15):4288-93.

Tsurusawa M, Ito M, Zha Z et al. Cell-cycle–associated expressions of proliferating cell nuclear antigen and Ki-67 reactive antigen of bone marrow blast cells in childhood acute leukemia. Leukemia 1992;6:669–674.

Jackie Walling, 'From methotrexate to pemetrexed and beyond. A review of the pharmacodynamic and clinical properties of antifolates', Investigational New Drugs 24: 37–77, 2006.

L.2 Supporting the prediction of the ALL Oncosimulator parameter values by using machine learning methods on gene expression data

L.2.1 Introduction

The Oncosimulator, as presented in the p-medicine deliverables D12.1, D12.2 and D12.3 as well as in the previous chapters of the present deliverable, is intended to be an important tool aiming at simulating the progression of tumours and their response to tumour type specific treatment in a personalized, patient-specific context. However, in order for these simulations to be able to reliably predict tumour growth and response to treatment for any given patient, a careful choice of the values of the Oncosimulator input parameters should be made prior to the execution of the simulation. A description of the types of input parameters is available in section L.1.4. Although such values could be inferred by calibrating the model, in the case where time course data for the progression of the disease and/or the response to treatment are available, choosing parameter values for a patient's case in which only the disease state for the day of diagnosis is known could be a difficult task. Although an extensive literature review based on the patient's clinical profile may conclude to specific guesses, such guesses may not be adequately accurate. On the other hand nowadays the number of patients for whom high throughput –omics (e.g. genomics, proteomics, metabolomics etc.) data is being made available gradually rise thanks to the advancement of high-throughput molecular screening techniques and the accompanying reduction of the cost for such experiments. Such types of data e.g. whole genome gene expression data can be very informative for the molecular status of the patient's tumour cells. This can lead to the prediction of their phenotypic characteristics at the cellular level which are related to the Oncosimulator input parameters. This implies that a correlation between the patient's molecular data and several Oncosimulator input parameters should exist.

For the case of the Acute Lymphoblastic Leukemia p-medicine Oncosimulator branch, the whole genome expression data (presented in detail in L.2.3.1), were made available for a number of patients together with their clinical and treatment response data. As presented in section L.1.4 patient specific estimates of the Cell Kill Rate (CKR) parameter have been inferred for the drug prednisone (PRED). In this sub-chapter, the first steps of a machine learning-based method capable of predicting a suitable value for the PRED CKR parameter are presented. The CKR value corresponds to a newly introduced patient and is based on his or her whole genome expression data. The relationship between the expression data and the CKR parameter has been extracted by fitting machine learning models. The estimated CKR values for the patients included in the training set (as estimated in section L.1.4) are used as responses and the patients' gene expression are used as predictors. It should be mentioned, however, that since the estimation of the CKR parameter has been made in a univariate context (i.e. the growth cell kinetics have been assumed identical for all patients, setting the majority of the remaining Oncosimulator parameters constant for the whole dataset), the response values (CKR estimates) have a potential to change when a multivariate approach of parameter estimation is to be followed in the subsequent steps of the analysis. The latter have been planned for the last year of the project. In contrast, the predictor values (gene expression data) are stable measurements for each given patient. Nevertheless, interesting conclusions could be extracted even from these preliminary steps of the present analysis and subsequently, the requirements for the more advanced phase of analysis could be formulated.

Finally, based on an extensive literature review concerning the acquisition of predictive gene signatures and the prediction of cellular phenotypes using gene expression data, methods that

aggregate the gene expression on a gene-set or pathway level have been selected to be exploited for reasons of robustness and interpretability presented in detail in the following section.

L.2.2 Pathway-based gene expression analysis

Since the advancement and the spreading of high-throughput molecular screening in in the late 90's and early 2000's, a plethora of high dimensional datasets have been made available both in the laboratory and at the clinical level. Among the many types of –omics data based experiments one of the most frequently conducted ones is the measurement of gene expression using microarrays that produce measurements at a genomic scale, usually for thousands of genes. Although these measurements have allowed the detailed statistical study of the transcriptome, identifying, for example, the differentially expressed genes between two groups of patients or between two cellular phenotypes, the need for biologically interpretable findings has led to the creation of a gene-set based analysis of gene expression data. A type of gene-sets - among several others - are the curated pathways such as those stored in the KEGG (http://www.genome.jp/kegg/pathway.html), the BIOCARTA (http://www.biocarta.com/) and the REACTOME (http://www.reactome.org/) databases.

Two widely accepted and frequently exploited approaches on the functional enrichment of gene expression data are the Over-representation analysis and the Gene-Set Analysis. In the former, differentially expressed genes are first identified between two groups in the data and subsequesntly the pathways (or in general the functional categories) that these genes participate in are found. In the latter, the differential expression statistics (such as the p-values of a t-test) of all genes included in the dataset are computed and each pathway (or gene-set) is analyzed in order to establish whether the genes that constitute it collectively show a significant differential expression. Two representative methods for these types of analysis are DAVID (Huang et al., 2007a, 2007b; Sherman et al., 2007) and GSEA (Subramanian et al., 2005) respectively. What is clear even from this short description of the two approaches is that in both approaches individual genes are first tested for differential expression between groups and the subsequent functional analysis for the identification of differentially regulated pathways refers to the differences between these groups.

An alternative approach that is gradually gaining interest is the one in which gene expression data is first aggregated at the pathway level (or in general at the gene-set level) in such a way that each row corresponds to a measurement of pathway activity, enrichment or expression. Following this transformation, the pathway-level data is directly analyzed in order for differentially expressed pathways to be identified. A significant difference between this approach and the previously presented methods is that the aggregation is made for every single sample of the dataset and so the transformed dataset may be used in a wider range of analysis tasks. This in conjunction with the identification of differentially expressed pathways may also be the classification or the clustering of samples using pathway-related features or even the regression analysis in the space of pathways.

The basic differences between the approaches are depicted in Fig. L.2.1:

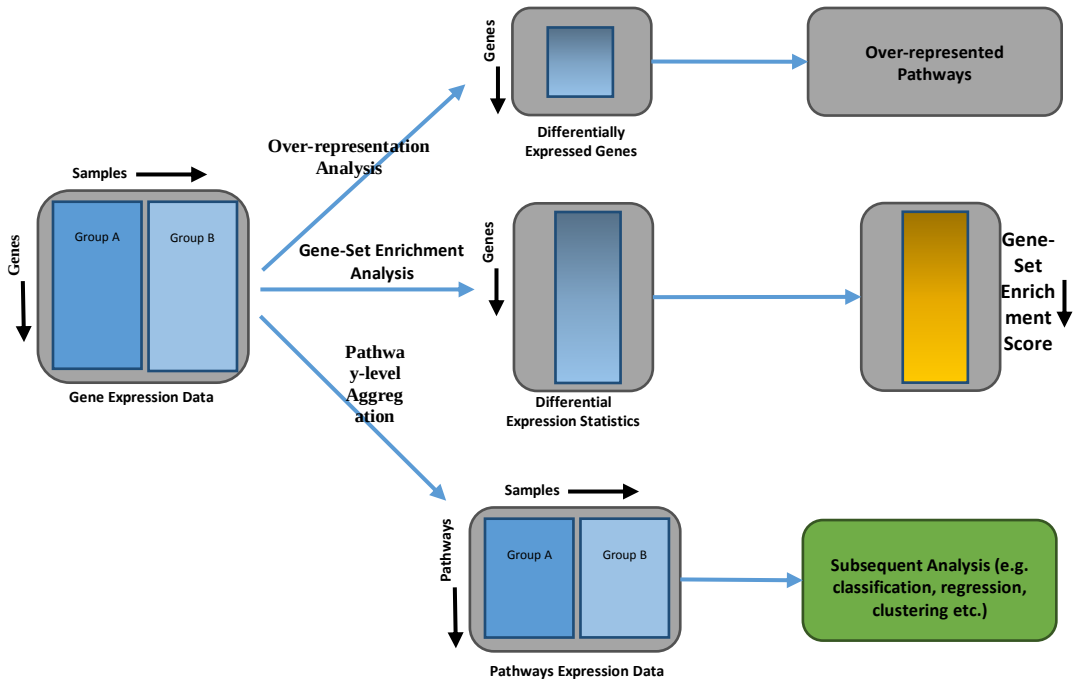


Fig. L.2.1 Diagrammatic description and comparison of Over-representation, Gene-Set Enrichment and Pathway-level Aggregation Analysis.

As stated in literature (Hwang, 2012a; Khatri et al., 2012; Varadan et al., 2012), in general the transformation of the gene expression data from the gene space to the pathway space is expected to lead to increased robustness of the results of the downstream analysis of molecular data in contrast with the gene signatures that are commonly found to be unstable. Moreover, the aggregation of gene expression to another commonly shared space i.e. the space of pathways is believed to be able to reduce the intrinsic technological and biological variances across sample. A number of well established methods for the aggregation of gene expression data to pathway activity/expression/enrichment in a single-sample context are given in (Barbie et al., 2009; Greenblum et al., 2011; Hänzelmann et al., 2013; Hwang, 2012b; Lee et al., 2008; Tomfohr et al., 2005.) Interesting reviews on the comparison of these kinds of methods are available in (Hwang, 2012a; Varadan et al., 2012).

L.2.3 Supporting the prediction of the ALL Oncosimulator cell kill rate parameter value

L.2.3.1 Description of the available data

In the context of p-medicine anonymized data describing the clinical characteristics and the disease progression for a number of ALL patients, presented in detail in section L.1.2 and coupled with whole genome expression data, have been made available. As described in section L.1.4 the Cell Kill Rate (CKR) parameter for the drug prednisone has been estimated for a number of 151 patients. Regression analysis has been performed on the available data. An additional number of patients have been excluded from the analysis of the present section because they have showed identical CKR estimates approximating the highest value (0.8) although they have shown differences in terms of gene expression. The latter may mislead the regression analysis. Finally, a number of 136 patient cases have been taken into account in the present analysis. However, as already stated in the

Introduction the present analysis is based on preliminary results of CKR estimation analysis and the estimation of these parameters for the full set of patients together with the subsequent regression analysis is to be refined during the last year of the project.

As far as gene expression data is concerned, the dataset consists of measurements from a number of ~40.000 probes and the experimental procedures followed are described in detail in (Cario et al., 2005). Finally, information about this gene expression data is available in the GSE4057 entry of Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo).

L.2.3.2 Data preprocessing, missing value imputation and pathway level aggregation

Before training any machine learning method with the available gene expressions and estimates of the CKR data, specific preprocessing steps have to be followed in order for the data to be prepared for the pathway-based gene expression analysis. Moreover, missing values are commonly found in high-throughput screening data so these values have to be filled in by imputation methods before any further step of analysis could be done.

L.2.3.2.1 Collapsing gene expression from the probe level to the gene level

The first preprocess step taken has been for every sample to collapse the measurements of gene expression in the probe level (in our dataset CloneIDs) to the Gene level (EntrezIDs). More specifically in the gene expression microarray platforms usually different probes that refer to the same gene are included and since the gene sets referring to pathways are collections of genes (usually annotated by EntrezIDs), the dataset should be transformed from the probe level to the gene level. In order for this step to be executed the collapseRows function(Miller et al., 2011) of WGCNA R (Langfelder and Horvath, 2008) package has been used. Among the different choices for the calculation of the expression at gene level, the max mean (choosing the probe among those referring to the same gene that shows the maximum mean expression), has been chosen to be executed as proposed in (Miller et al., 2011) for the type of downstream analysis to be pewrformed. Moreover, it should be stressed that this preprocessing procedure may lead to a dataset with fewer missing values since the probes with the fewest missing values may be asked to be selected by the CollapseRows function.

This preprocessing step, apart for rendering the gene expression datasets compatible for pathway-based analysis, may also strengthen their biological interpretability and more importantly may promote the simultaneous analysis of data measured by different microarray platforms (Miller et al., 2011). This has also been the case for the available dataset where slightly different platforms have been used for certain groups of patients (see GSE4057 entry of Gene Expression Omnibus database).

In order for the collapsing procedure to be implemented, a mapping between CloneIDs accession numbers and EntrezIDs should be provided in the CollapseRows function. This mapping file has been created using the online conversion platform of Stanford Microarray Database (http://smd.princeton.edu/cgi-bin/source/sourceBatchSearch).

After excluding the probes that do not map to any EntezID and by collapsing the remaining probes by their mapping to EntrezIDs, we have concluded with a dataset that includes the expression levels for 14227 genes.

L.2.3.2.2 Gene filtering

Although the step of collapsing probes to genes, as described in the previous section, may lower the number of missing expression values in the datasets, a significant number of these types of values

may still exist. Before trying to estimate the missing values using the non-missing values of the dataset (as is described in detail in the following section), the genes that still show a high frequency of missing values (>15%) across the dataset have been filtered out. Those genes may be thought to have been inadequately measured. Additionally the performance of the imputation algorithms in terms of accuracy generally drops significantly for datasets with missing value frequency higher than 15% (Brock et al., 2008; Luengo et al., 2012).

The filtering step has been implemented using the goodGenes function of the WGCNA package in R(Langfelder and Horvath, 2008) by selecting the minimum fraction of non-missing samples for a gene to be considered "good" to be 85%. The result of this step has been the exclusion of 5464 genes that were showing many missing samples or zero variance.

L.2.3.2.3 Missing value imputation

Following the exclusion of a significant number of missing values, the remaining unsuccessfully measured values of gene expression were filled by imputation. This step is believed to be preferable to the exclusion of all genes that show missing values, since the experience accumulated in literature (Aittokallio, 2010; Brock et al., 2008; Jörnsten et al., 2005; Liew et al., 2011; Luengo et al., 2012) has shown that an improper handling of missing data may hinder an effective downstream analysis.

The algorithm selected in order for the missing values to be imputed has been the imputation by the Nearest Neighbors algorithm (knn-imputation) (Troyanskaya et al., 2001.) This is one of the simplest algorithms of this kind but it is widely and frequently used. Although more sophisticated algorithms do exist (like Bayesian Principal Component Analysis (Oba et al., 2003), Local Least Square impute (Kim et al., 2005) and Least Square Adaptive (Bø et al., 2004)), which are shown to be capable of producing better results[(Brock et al., 2008; Luengo et al., 2012)], the computational cost for these algorithms is significantly higher compared to knn (for example BPCA needs about 20 hours to be completed in a typical desktop computer setting). Therefore, during these preliminary steps of the analysis it has been decided to select a more lightweight algorithm letting the exploitation of the more powerful algorithms for the subsequent more advanced steps of the analysis. The knn imputation has been implemented by using the Impute package in Bioconductor-R and by choosing the parameter k (the neighbors that are used in order for a missing value to be imputed) to be 20. The latter is one of the most commonly chosen values for this parameter[(Brock et al., 2008)].

L.2.3.2.4 Gene-set/pathway level gene expression aggregation

Possessing now a gene expression dataset at the gene level without missing values, the gene-set/pathway-based analysis can be implemented. The underlying theory has been presented in section L.2.2. The method selected to be used has been the Gene Set Variation Analysis (GSVA)(Hänzelmann et al., 2013) that utilizes the associate Bioconductor-R package GSVA. The gene sets chosen in order to aggregate the gene expression values have been those referring to genes constituting the KEGG pathways. There are 186 of them. The gene sets (CP:KEGG: KEGG gene sets) have been download from the Broad Institute MsigDB (http://www.broadinstitute.org/gsea/msigdb) as a *.gmt* file and have been introduced into R using the getGmt function of the GSEABase package. It should be mentioned, however, that in the subsequent steps of the analysis, additional gene-sets have been planned to be exploited in order to study the influence of the gene set selection on the accuracy of the final predictions. Following the application of the GSVA method, the dataset now consists of 186 features (enrichment score for KEGG pathways) for every one of the 136 samples.

L.2.3.3 Training the machine learning model and cross-validation results

L.2.3.3.1 Problem definition and machine learning algorithms tested

In order for a machine learning problem to be defined the characteristics of the available dataset - as far as the included variables are concerned - should be examined. Following the previously described preprocessing and transformation steps, the dataset consists now of 186 numerical features that correspond to the activity of KEGG pathways in a single sample context (the predictors of the dataset) and one numerical value (defined in the range (0,1)) that is the estimated value of the CKR parameter for every single sample/patient. Therefore, in order to acquire a model that would be able to predict a reliable CKR parameter value for a newly introduced sample/patient based on pathway activity values, a regression analysis should be performed and a regression model should be finally created.

The machine learning model inference algorithms chosen to be tested are Support Vector Machines (Cortes and Vapnik, 1995) (in three variations concerning the choice of the kernel which could be Linear, Polynomial or a Radial Basis Function) and Random Forests (Breiman, 2001). These two algorithms are two of the most widely used machine learning ones that are capable of inferring both classification and regression models and can learn non-linear relationships that may exist in the dataset.

In order for the tests to be executed, the caret Package in R (http://caret.r-forge.r-project.org/) has been used. It contains numerous functions to streamline the model training process for regression and classification problems and incorporates machine learning algorithms from 21 existing R packages. Moreover, the caret Package offers features that support the evaluation of the effect of the model tuning parameters (the parameters of the machine learning algorithms) on performance, the selection of the "optimal" model across these parameters and the estimation of the model performance from a dataset offering various types of cross-validation.

L.2.3.3.2 Cross-validation results using the entire feature set

Although it is a common practice in studies trying to create predictive models based on high dimensional molecular data to use only a subset of features - for example the significantly differentially expressed genes between two phenotypes - here, it has been decided to train the machine learning models by the entire feature set. The reason for this has been a massive reduction in the dimensionality of the problem obtained by aggregating the gene expression measurements to pathway activity. It should be mentioned, however, that the Random Forests algorithm, by definition, has a built-in feature selection step (Breiman, 2001). The results of the procedures of training and evaluating, by a 10x10 cross validation, the models are available in Table L.2 .

Table L.2. Results of the cross-validated evaluation of the trained regression model performance on predicting the CKR parameter values based on gene expression data

Method	Sub-method	Tuning parameter			RMSE (10x10 cross validation)
Random Forests		**mtry** (number of predictors sampled for splitting at each tree node)			
		2			0.1507686
		48			**0.1507560**
		94			0.1513661
		140			0.1523421
		186			0.1524977
SVM					
	Linear Kernel	**C**			
		1			0.2346681
	Polynomial Kernel	**Degree**	**Scale**	**C**	
		1	0.001	0.25	0.1541765
		1	0.001	0.5	0.1522134
		1	0.001	1	0.1504459
		1	0.01	0.25	0.1498342
		1	0.01	0.5	0.1501210
		1	0.01	1	0.1564672
		1	0.1	0.25	0.1687962
		1	0.1	0.5	0.1879542
		1	0.1	1	0.2128265
		2	0.001	0.25	0.1523583
		2	0.001	0.5	0.1507105
		2	0.001	1	0.1503095
		2	0.01	0.25	0.1512965
		2	0.01	0.5	0.1518911
		2	0.01	1	0.1520950
		2	0.1	0.25	0.1503484
		2	0.1	0.5	0.1503484
		2	0.1	1	0.1503484
		3	0.001	0.25	0.1511744
		3	0.001	0.5	0.1505853
		3	0.001	1	0.1505792
		3	**0.01**	**0.25**	**0.1498330**
		3	**0.01**	**0.5**	**0.1498330**
		3	**0.01**	**1**	**0.1498330**
		3	0.1	0.25	0.1578186
		3	0.1	0.5	0.1578186
		3	0.1	1	0.1578186
	Radial Basis Kernel	**Sigma**		**C**	
		0.00325094		0.25	0.1516329
		0.00325094		0.50	0.1492191
		0.00325094		**1.00**	**0.1491777**
		0.00325094		2.00	0.1497389
		0.00325094		4.00	0.1497125

As can be seen in Table L.2, except for the SVM with a linear kernel case, the performance of the methods are comparable and the Root Mean Square error is approximately equal to 0.15. This means that, as estimated by the Cross-Validation, for a newly presented patient the estimation of the CKR parameter value that would be the most appropriate in order for the simulation of the Oncosimulator to give results close to reality, could be given with an accuracy of about 85%. Equivalently the predicted CKR value is expected to differ from the most appropriate value by at most 15%.

L.2.4 Conclusions

In the analysis presented, a regression model able to predict the most appropriate value for the Cell Kill Rate (CKR) parameter of the Oncosimulator (in order for the simulation of the model to converge to the observed response of a given patient to prednisone treatment) has been trained using gene expression data and the estimates for the CKR parameter value inferred by fitting the Oncosimulator on the patient's response to treatment data. As inferred by the Cross-Validation training and the evaluation of the tested models, the value for the CKR parameter could be predicted by only observing the gene expression data measured at the stage of diagnosis with an accuracy of 85%. Although such a deviation (15%) could significantly influence the behaviour of the Oncosimulator, it can be characterized as acceptable, since both the estimation of the CKR parameter value and the regression analysis are still in their preliminary stages.

Nevertheles, irrespective of the percentage of error, any narrowing of the range of possible values for a model parameter could be beneficial for the prediction of the proper parameter values for a newly introduced patient for whom only the status at diagnosis is known. It could also be beneficial for the procedure of more generic parameter estimation.

L.2.5 REFERENCES

Aittokallio, T., 2010. Dealing with missing values in large-scale studies: microarray data imputation and beyond. Brief. Bioinform. 11, 253–264.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., Fröhling, S., Chan, E.M., Sos, M.L., Michel, K., Mermel, C., Silver, S.J., Weir, B.A., Reiling, J.H., Sheng, Q., Gupta, P.B., Wadlow, R.C., Le, H., Hoersch, S., Wittner, B.S., Ramaswamy, S., Livingston, D.M., Sabatini, D.M., Meyerson, M., Thomas, R.K., Lander, E.S., Mesirov, J.P., Root, D.E., Gilliland, D.G., Jacks, T., Hahn, W.C., 2009. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 462, 108–112.

Bø, T.H., Dysvik, B., Jonassen, I., 2004. LSimpute: accurate estimation of missing values in microarray data with least squares methods. Nucleic Acids Res. 32, e34.

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.

Brock, G.N., Shaffer, J.R., Blakesley, R.E., Lotz, M.J., Tseng, G.C., 2008. Which missing value imputation method to use in expression profiles: a comparative study and two selection schemes. BMC Bioinformatics 9, 12.

Cario, G., Stanulla, M., Fine, B.M., Teuffel, O., Neuhoff, N.V., Schrauder, A., Flohr, T., Schäfer, B.W., Bartram, C.R., Welte, K., Schlegelberger, B., Schrappe, M., 2005. Distinct gene expression profiles determine molecular treatment response in childhood acute lymphoblastic leukemia. Blood 105, 821–826.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20, 273–297.

Greenblum, S.I., Efroni, S., Schaefer, C.F., Buetow, K.H., 2011. The PathOlogist: an automated tool for pathway-centric analysis. BMC Bioinformatics 12, 133.

Hänzelmann, S., Castelo, R., Guinney, J., 2013. GSVA: gene set variation analysis for microarray and RNA-seq data. BMC Bioinformatics 14, 7.

Huang, D.W., Sherman, B.T., Tan, Q., Collins, J.R., Alvord, W.G., Roayaei, J., Stephens, R., Baseler, M.W., Lane, H.C., Lempicki, R.A., 2007a. The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists. Genome Biol. 8, R183.

Huang, D.W., Sherman, B.T., Tan, Q., Kir, J., Liu, D., Bryant, D., Guo, Y., Stephens, R., Baseler, M.W., Lane, H.C., Lempicki, R.A., 2007b. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res. 35, W169–W175.

Hwang, S., 2012a. Comparison and evaluation of pathway-level aggregation methods of gene expression data. BMC Genomics 13, S26.

Hwang, S., 2012b. Comparison and evaluation of pathway-level aggregation methods of gene expression data. BMC Genomics 13, S26.

Jörnsten, R., Wang, H.-Y., Welsh, W.J., Ouyang, M., 2005. DNA microarray data imputation and significance analysis of differential expression. Bioinformatics 21, 4155–4161.

Khatri, P., Sirota, M., Butte, A.J., 2012. Ten Years of Pathway Analysis: Current Approaches and Outstanding Challenges. PLoS Comput Biol 8, e1002375.

Kim, H., Golub, G.H., Park, H., 2005. Missing value estimation for DNA microarray gene expression data: local least squares imputation. Bioinformatics 21, 187–198.

Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics 9, 559.

Lee, E., Chuang, H.-Y., Kim, J.-W., Ideker, T., Lee, D., 2008. Inferring Pathway Activity toward Precise Disease Classification. PLoS Comput Biol 4, e1000217.

Liew, A.W.-C., Law, N.-F., Yan, H., 2011. Missing value imputation for gene expression data: computational techniques to recover missing data from available information. Brief. Bioinform. 12, 498–513.

Luengo, J., García, S., Herrera, F., 2012. On the choice of the best imputation methods for missing values considering three groups of classification methods. Knowl. Inf. Syst. 32, 77–108.

Miller, J.A., Cai, C., Langfelder, P., Geschwind, D.H., Kurian, S.M., Salomon, D.R., Horvath, S., 2011. Strategies for aggregating gene expression data: The collapseRows R function. BMC Bioinformatics 12, 322.

Oba, S., Sato, M., Takemasa, I., Monden, M., Matsubara, K., Ishii, S., 2003. A Bayesian missing value estimation method for gene expression profile data. Bioinforma. Oxf. Engl. 19, 2088–2096.

Sherman, B.T., Huang, D.W., Tan, Q., Guo, Y., Bour, S., Liu, D., Stephens, R., Baseler, M.W., Lane, H.C., Lempicki, R.A., 2007. DAVID Knowledgebase: a gene-centered database integrating heterogeneous gene annotation resources to facilitate high-throughput gene functional analysis. BMC Bioinformatics 8, 426.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., Mesirov, J.P., 2005. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. U. S. A. 102, 15545–15550.

Tomfohr, J., Lu, J., Kepler, T.B., 2005. Pathway level analysis of gene expression using singular value decomposition. BMC Bioinformatics 6, 225.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., Altman, R.B., 2001. Missing value estimation methods for DNA microarrays. Bioinformatics 17, 520–525.

Varadan, V., Mittal, P., Vaske, C.J., Benz, S.C., 2012. The Integration of Biological Pathway Knowledge in Cancer Genomics: A review of existing computational approaches. IEEE Signal Process. Mag. 29, 35–50.

C Overall conclusions

In the present document the crucial steps of the initial adaptation and optimization of the models constituting the three branches of the p-medicine Oncosimulator i.e. the nephroblastoma (N) branch, the breast cancer (B) branch and the acute lymphoblastic leukemia (L) branch have been outlined. The N and L branches have been clinically adapted whereas the B branch has been adapted to pertinent animal data at this stage. The latter was due to delays of legal nature that have been encountered in the data transmission procedure. Despite particular limitations applicable to different tumour types, the overall outcome of the procedure is positive and encouraging. It has demonstrated the potential of the various branches of the Oncosimulator to reasonably simulate clinical or experimental animal reality in a consistent and rational way. Obviously the more data is utilized, the better the clinical adaptation is expected to become. Therefore, additional adaptation and partial validation efforts are to be undertaken in the remaining project lifetime.

Appendix A - Abbreviations and Acronyms

ALL	Acute Lymphoblastic Leukemia
CAU	Christian-Albrechts-Universität zu Kiel
CKR	Cell kill rate
CTL	Cytotoxic T Lymphocytes
HR	High Risk
ICCS-NTUA	Institute of Communication and Computer Systems – National Technical University of Athens
ISOG	*In Silico* Oncology Group
LIMP	LImited Mitotic Potential (cell)
LUH	Leibnitz Universitaet Hannover
MRD	Minimal Residual Disease
SR	Small Risk
TC	Tumour Cell
USAAR	University of Saarland
WP	workpackage
WT	Wilms Tumour (nephroblastoma)